CONNECTIONS

Official Journal of the International Network for Social Network Analysis

INTERNATIONAL NETWORK FOR SOCIAL NETWORK ANALYSIS

CONNECTIONS is the official journal of the **International Network for Social Network Analysis** (INSNA). INSNA is a scientific organization made up of scholars across the world. Updated information about the INSNA's annual conference (**Sunbelt Social Network Conferences**) can be found on the website at *www.insna.org*.

INSNA includes official board members and five committees:

BOARD MEMBERS

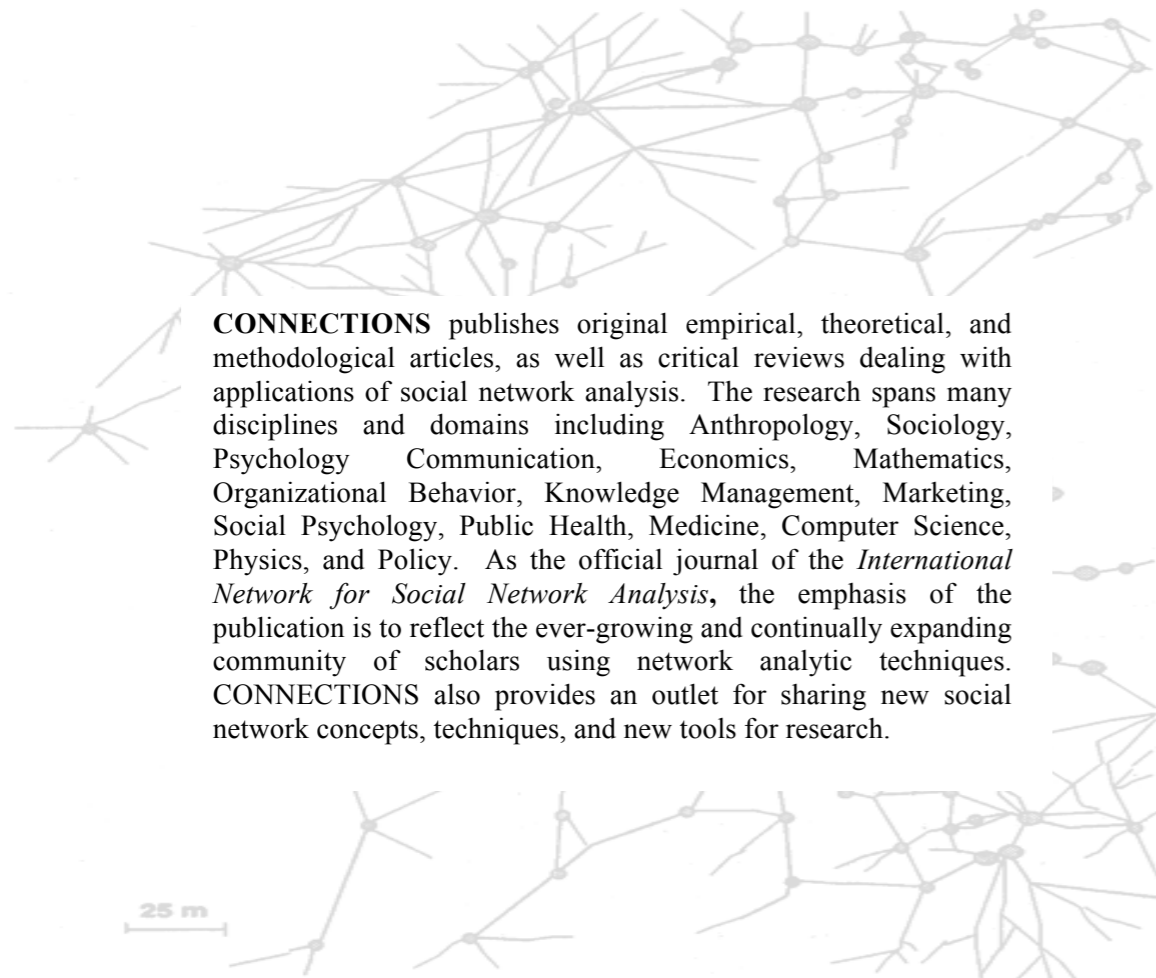|	
President:	George A. Barnett
Vice President:	Pip Pattison
Treasurer:	Thomas W Valente
Founder:	Barry Wellman
Members:	Phil Bonacich, Martin Everett, Katie Faust, Scott Feld, Anuska Ferligoj, Garry Robins, Ulrik Brandes, David Lazer

COMMITEES

Finance Committee **-** Chaired by Treasurer Tom Valente

Conference Committee **-** Chaired by Sunbelt host Mario Diani

Web Committee **-** Chaired by Chief Information Officer Benjamin Elbirt

Publications Committee - Chair – TBD

Composed of current and former editors of *Social Networks*, *CONNECTIONS* and *Journal of Social Structure* (JOSS) to oversee the INSNA's relations with the publications, selection of *CONNECTIONS* 'and *JOSS's* future editors and to coordinate the publications so that they are complementary rather than in competition with one another. To insure openness to new ideas, one or more additional members will be selected by the President.

CONNECTIONS publishes original empirical, theoretical, and methodological articles, as well as critical reviews dealing with applications of social network analysis. The research spans many disciplines and domains including Anthropology, Sociology, Psychology Communication, Economics, Mathematics, Organizational Behavior, Knowledge Management, Marketing, Social Psychology, Public Health, Medicine, Computer Science, Physics, and Policy. As the official journal of the *International Network for Social Network Analysis*, the emphasis of the publication is to reflect the ever-growing and continually expanding community of scholars using network analytic techniques. CONNECTIONS also provides an outlet for sharing new social network concepts, techniques, and new tools for research.

Front Cover: *The four figures are from enclosed article titled "Co-authorship in Italian Workshops on Population Studies: An Analysis with a Network Approach" by Giulia Rivellini and Laura Terzera. Images display collaboration networks observed cross-sectionally for each of the four national conferences on population studies (Giornate di Studio sulla Popolazione). The nodes are authors, while the relationship that links them together is their collaboration in writing and/or preparing a presentation. The network is observed by gender (shape), affiliation (color), number of contributions (size of nodes), and number of collaborations inside the same research group (line thickness).*

Dr. Thomas Valente, Editor
Professor, Director of the Master of Public Health Program
Professor in the Department of Preventive Medicine,
University of Southern California, Alhambra, CA
tvalente@usc.edu

Dr. Kathryn Coronges, Managing Editor
Assistant Professor, Department of Behavioral Sciences & Leadership
United States Military Academy, West Point, NY
kcoronges@gmail.com

Joseph Dunn, Associate Editor
Garrison, New York, *josephdunn9@gmail.com*

Editorial Headquarters
University of Southern California, Institute of Prevention Research
1000 Fremont Ave., Unit #8, Building A, Room 5133, Alhambra, CA 91803
Tel: (626) 457-4139; fax: (626) 457-6699

International Network for Social Network Analysis
Hardcopy circulation of Connections is sent to all members of INSNA, the International Network for Social Network Analysis, which has reached just over one thousand members. Subscription to CONNECTIONS can be obtained by registering for INSNA membership through the website*: www.insna.org.* Standard membership fee is US$60 ($40 for students). Wherever possible, items referenced in articles (such as data and software) are made available electronically through the INSNA website. In addition, the website provides access to a directory of members' email addresses, network datasets, software programs, and other items that lend themselves to electronic storage.

Sunbelt Social Network Conferences
Annual conferences for INSNA members take place in the United States for two years and in Europe every third year. The Sunbelt Conferences bring researchers together from all over the world to share current theoretical, empirical and methodological findings around social networks. Information on the annual Sunbelt Social Network Conferences can also be found on the INSNA website. Sunbelt XXX will be held in Riva del Garda, Italy from June 29-July 4, 2010.

Letters to the Editor
As a new feature of CONNECTIONS, we welcome letters to the Editor on all issues of academic interest. Letters should be brief and will be subject to editing and condensation. In addition, book reviews, network images or review articles will be considered for publication.

INSTRUCTIONS TO AUTHORS

CONNECTIONS publishes original empirical, theoretical, tutorial, and methodological articles that use social network analysis. The journal publishes significant work from any domain that is relevant to social network applications and methods. Commentaries or short papers in response to previous articles published in the journal are considered for publication. Review articles that critically synthesize a body of published research are also considered, but normally are included by invitation only. Authors who wish to submit a commentary, book review, network image or review article are welcome to do so.

Submitting Manuscripts
Authors are required to submit manuscripts online to the editor, Thomas Valente at *tvalente@usc.edu*. Expect a notice of receipt of your manuscript via email within one week. Feedback from the editor and reviewers will be sent to the corresponding author within six months after receipt. Revised or resubmitted manuscripts should include a detailed explanation of how the author has dealt with each of the reviewer's and Editor's comments. For questions or concerns about the submission process, authors should contact the editor.

Manuscripts must be in MS Word format and should not exceed 40 pages including tables, figures and references. Manuscripts should be arranged in the following order: title page, abstract, corresponding author contact information, acknowledgments, text, references, and appendices. Abstracts are limited to 250 words. Tables should be in Word or Excel format and embedded in the document. Format and style of manuscript and references should conform to the conventions specified in the latest edition of Publication Manual of the *American Psychological Association*. Please disable any automatic formatting when possible. A figure and its legend should be sufficiently informative that the results can be understood without reference to the text. In each issue, we select an image from an accepted article to appear on the front cover of the journal.

CONNECTIONS

ARTICLES

Leadership Insularity:
A New Measure of Connectivity Between Central Nodes in Networks

Samuel Arbesman
Department of Health Care Policy, Harvard Medical School
Cambridge, Massachusetts

Nicholas A. Christakis
Departments of Health Care Policy & Medicine, Harvard Medical School
Department of Sociology, Harvard University
Cambridge, Massachusetts

We combine two foci of interest with respect to community identification and node centrality and create a novel metric termed "leadership insularity." By determining the most highly connected nodes within each community of a network, we designate the 'community leaders' within the graph. In doing this, we have the basis for a novel metric that examines how connected, or disconnected, the leaders are to each other. This measure has a number of appealing measurement properties and provides a new way of understanding how network structure can affect its dynamics, especially information flow. We explore leadership insularity in a variety of networks.

Authors

Samuel Arbesman, Ph.D. (Harvard University) is a posdoctoral research fellow in the Department of Health Care Policy at Harvard Medical School and affiliated with the Institute for Quantitative Social Science at Harvard University.

Nicholas A. Christakis, M.D., Ph.D., M.P.H (Harvard University) is Professor of Medical Sociology in the Department of Health Care Policy at Harvard Medical School; Professor of Medicine in the Department of Medicine at Harvard Medical School; and Professor of Sociology in the Department of Sociology in the Harvard Faculty of Arts and Sciences.

Correspondence: *Contact Samuel Arbesman at Harvard Medical School, Department of Health Care Policy, 180 Longwood Avenue, Boston, MA 02115. Phone: 617-432-7421; fax: 617-432-5891; email: arbesman@hcp.med.harvard.edu.*

INTRODUCTION

In recent years, there has been considerable work in two areas of network measurement: community identification and node centrality. Communities within networks are often identified as subgraphs that are connected more tightly than the graph as a whole. The available algorithms vary widely and include traditional clustering techniques, centrality-based community detection, and modularity-based methods (Porter, Onnela, & Mucha, 2009). Furthermore, there are many methods of determining the most centrally located nodes within a network. These range from examining the node with the highest degree to the node with the highest betweenness centrality and so forth (Newman, 2003).

Here, we combine these methods and create a novel metric known as "leadership insularity." By determining the most highly connected nodes within each community of a network, we are able to determine the 'community leaders' within the graph. In doing this, we have the basis for a novel metric that examines how connected, or disconnected, the leaders are connected to *each other*. This measure can be used to characterize individual leaders in a network (in terms of how isolated they are from other leaders) or it can be used to summarize the property of a whole network (in terms of how isolated its leaders are compared to other networks). This measure of insulation provides a new way of understanding how network structure can affect its dynamics, especially information flow.

Using a topographic analogy, as in Figure 1, each community may be viewed as an individual mountain within a mountain range, with its leader as the peak. The topography of the mountain range can vary wildly, and has implications for how closely connected the peaks are. Analogously, if the 'slope' of a community were shallow, two leaders would only be able to interact via many intermediaries. However, if the distance is much closer, then they might be able to interact more effectively. This has implications for many situations, such as coordination problems (Kearns, Suri, & Montfort, 2006).

Guimera et al. hint at something similar to leadership insularity, though their metrics are somewhat different (Guimerà, Mossa, Turtschi, & Amaral, 2005). They identify a number of different categories of nodes and even create a metric called the participation coefficient (which examines how connected nodes in a community are connected to other communities). Our measure is different in that it is mathematically simpler, by focusing only on the leaders of the communities, as opposed to all nodes. Moreover, since community leaders often have an outsized influence on the dynamics of their groups, it is useful to have a *single* metric for an entire network's leadership insularity.

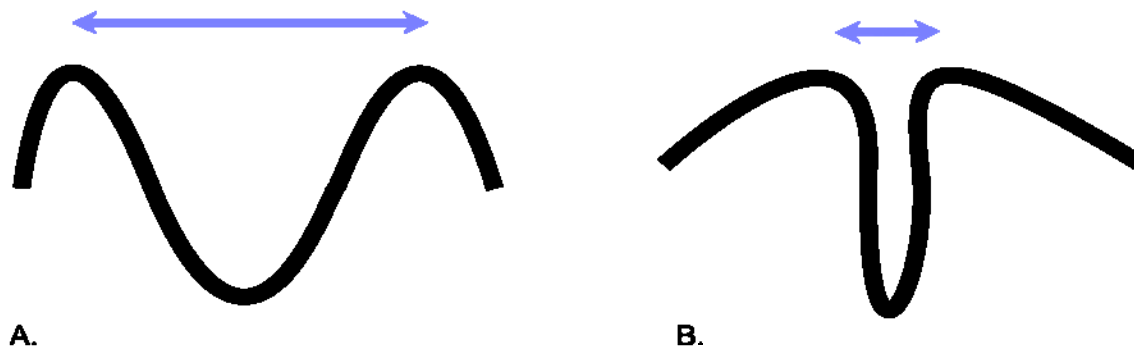**Figure 1. A Metaphorical View of Leadership Insularity**

Figure 1. Using the topographical imagery provided in the text: Part A has a large distance between leaders/peaks, while Part B has a much smaller distance between leaders.

By being able to quantify the distance between these community leaders, we can understand the structure and dynamics of networks better. After explaining the metric, which has some appealing measurement properties, we explore the leadership insularity of a variety of networks and examine how it relates to the diverse functions of these networks.

METHODS

1. Description of the Metric

Leadership Insularity is simply defined as the average relative distance between the leaders of different communities. This is achieved by dividing the path length between each leader by the average path length between any two individuals of their respective communities. The overall leadership insularity then becomes the average of these relative path lengths, weighted according to the size of the communities. The equation, visualized in Figure 2, is as follows:

$$I = \frac{1}{(N_c - 1)N} \sum_{i=1}^{N_c} \sum_{j=i+1}^{N_c} \frac{d(L_i, L_j)}{d(i,j)} (N_i + N_j)$$

$$(1)$$

Where the variables are defined as follows:

N_c = number of communities identified
N = number of nodes in the network
N_i = number of nodes in community i
L_i = leader of community i
$d(L_i, L_j)$ = distance between community leaders L_i and L_j
$d(i,j)$ = mean distance between communities i and j

The term:

$$\frac{1}{(N_c - 1)N} \sum_{i=1}^{N_c} \sum_{j=i+1}^{N_c} (N_i + N_j) \text{ equals 1 and}$$

is used to allow a weighted average of the various relative distances between community leaders.

In addition, the leadership insularity can be calculated for a single leader within the network as follows:

$$I_i = \frac{1}{2N_c N} \sum_{j \neq i}^{N_c} \frac{d(L_i, L_j)}{d(i,j)} (N_i + N_j)$$

$$(2)$$

When the mean of these individual leadership insularities is taken, the leadership insularity of the entire network is obtained.

The communities can be identified by a variety of methods, as can the community leaders. For the purposes of the implementation of the metric, we used the method described in Clauset to identify communities within our networks (Clauset, 2005). The community leaders were those nodes with the highest betweenness centrality when a community was viewed as a graph, separate from the network as a whole. If there are two or more nodes with equally high betweenness centralities, then a comparison is made to the nodes with the highest degree centrality. A randomly selected node from the intersection of the nodes with the highest betweenness and degree centralities is chosen (if the intersection has no nodes, then a randomly selected node from the highest betweenness centralities is used). This use of a combination of centrality measures is similar to that used by researchers studying peer-education and food intake (D. Buller et al., 2000; D. B. Buller et al., 1999).
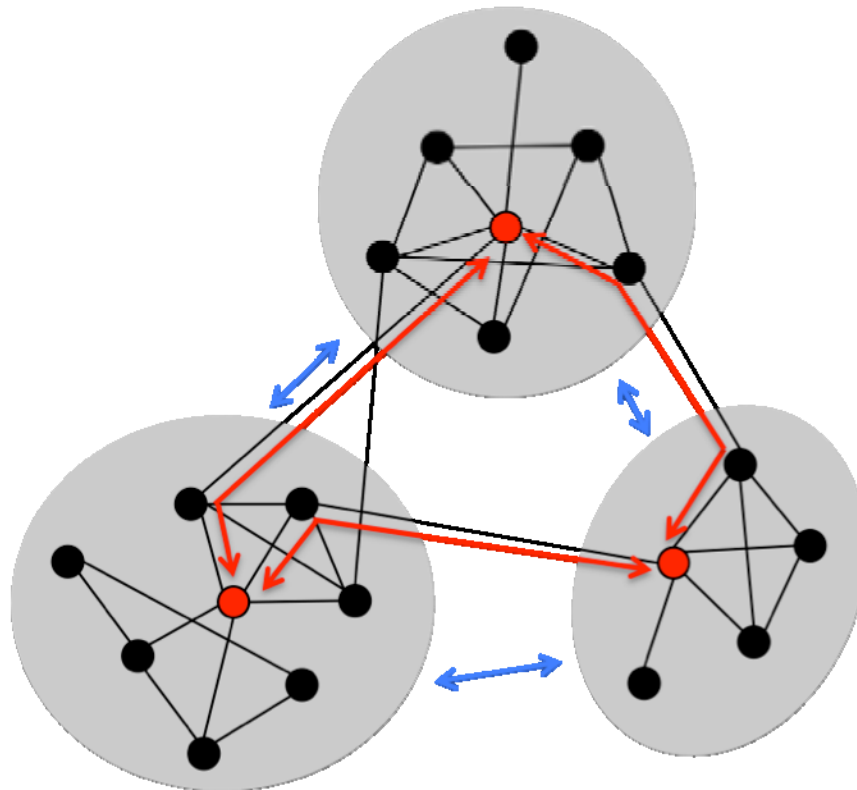
Figure 2. Visual Demonstration of Leadership Insularity

Figure 2. Red nodes indicate community leaders and red lines indicate distance between them. Blue lines indicate the mean distance of the individuals in one community to another. The calculated leadership insularity is 0.68. Figure adapted from Newman (Newman, 2003).

In the Addhealth dataset the number of communities with multiple equally good choices as leader is 3.2% of the total 1570 communities within the networks, with the majority of these only containing two possible leaders, and most of these possible leaders being the most central nodes for both measures (see section 2A). However, it seems that these numbers might be domain-specific. For example, one of our scientific collaboration datasets. Condensed Matter arXiv 2003 (see section 2B), had multiple equally good choices for the leader in about 25% of the communities, and these communities contained more than two possible leader choices (often around seven). Therefore, leader identification in different domains merits further study.

In addition, we performed a robustness test on the use of betweenness centrality for leader detections by creating a modified metric that uses degree centrality as the primary criterion (with betweenness centrality as the secondary criterion). Using this modified metric, a similar dispersion of leadership insularity was found in the Addhealth dataset as below, and similar correlations (albeit with less significant p-values).

The code has been implemented in Python and requires the packages of igraph and NetworkX (Csárdi & Nepusz, 2006; Hagberg, Schult, & Swart, 2008). It is being released under the GPL license and will be downloadable from the following locations:

http://christakis.med.harvard.edu/

http://arbesman.net/

Figure 3. Dispersion of Leadership Insularity in Schools
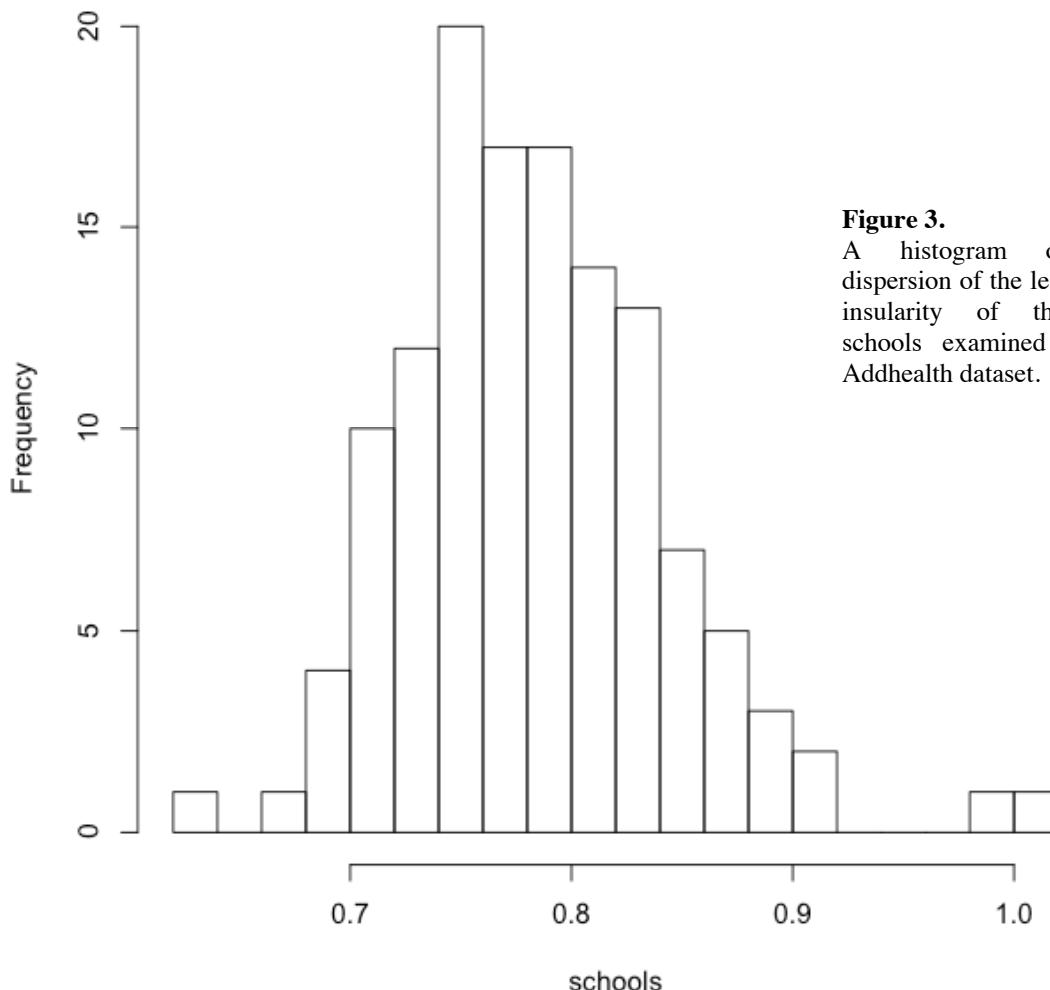
Figure 3.
A histogram of the dispersion of the leadership insularity of the 142 schools examined in the Addhealth dataset.

2. Applications

2.1 Addhealth Dataset

To test the robustness and applicability of leadership insularity, we applied the metric to a variety of networks. Our first test consisted of examining high school social networks in different schools. We expected that there would be significant variation between schools, and that this variation would be related to other differences between schools. We used the Addhealth dataset, a survey conducted in 142 American high schools (Harris, 2008). As part of the survey, adolescents were asked about their

social ties, which allowed us to reconstruct the social networks for each high school.

A high degree of dispersion was found in the high schools, as seen in Figure 3. In addition, we observed a significant relationship between a high school's leadership insularity and certain other attributes of the schools, such as the extent to which students feel safe at school or the average tenure of the students in the school. For example, a simple OLS regression model reveals that schools with a high LI had a higher duration of time the students had been in the school, regression coefficient = 6.69, $p < 0.0001$ (standard error = 1.37). Schools with high LI also had students who were more likely to report

feeling safe in the school, regression coefficient = 1.69, p=0.003 (standard error = 0.555). The longer the average duration of the students in a school could very easily lead to a certain amount of social insularity, which would in turn lead to leadership insularity. Less turnover in the nodes on the network also would stabilize the cliques in the schools, and their leaders. Similarly, this type of social insularity might lead to a greater feeling of safety in one's school and neighborhood, since one's social circle is cloistered and insulated from the world at large.

Scientific Coauthorship Networks. We also examined the variation in leadership insularity for various scientific coauthorship networks. These networks are constructed from authorship of scientific papers, where two individuals are connected if they coauthored a paper. We examined the coauthorship networks compiled from selected subareas within arXiv, an online preprint repository with a physics focus. The areas we looked at are theoretical high energy physics (hep-th), condensed matter (cond-mat), and astrophysics (astro-ph) (Newman, 2001). In addition, a smaller dataset composed network science articles (netscience) was also included (Newman, 2006). As a check, we also used a more recent version of the condensed matter coauthorship network (up to 2003, as opposed to 1999) to ensure that each area's leadership insularity was reasonably robust.

As seen in Table 2, there is a certain amount of variation in the leadership insularities of the different scientific disciplines. This could be due to a variety of factors, such as the degree of collaboration within the networks. Patterns of collaboration and interaction vary between scientific areas, and these differences are visible in differences between leadership insularity. In addition, with more data available, such as the number of citations (as an indication of the impact of the discipline), it could be seen whether or not the connectivity between scientific 'leaders' has an impact on the productivity of a discipline or leader.

Table 2. Leadership Insularity of Scientific Subdisciplines

arXiv Area	Leadership Insularity
hep-th	0.70
netscience	0.69
cond-mat	0.77
astro-ph	0.76
cond-mat-2003	0.76

CONCLUSIONS

Large groups configured as networks have subgroups, and subgroups typically have leaders. The ability of the group as a whole to function may be related to how integrated its leaders are with each other, and not just with their own group members, especially when communication flows between leaders are indirect (through others) and not direct (in the form of person-to-person ties). Otherwise similar networks may therefore differ meaningfully in terms of how inter-connected their leaders are, and this measure may correlate with a variety of internal and external properties of the network. We have proposed a novel metric, termed leadership insularity, to capture the degree of social isolation of central nodes of different communities within networks.

REFERENCES

Buller, D., Buller, M. K., Larkey, L., Sennott-Miller, L., Taren, D., Aickin, M., et al. (2000). Implementing a 5-a-Day Peer Health Educator Program for Public Sector Labor and Trades Employees. *Health Education & Behavior, 27*(2), 232-240.

Buller, D. B., Morrill, C., Taren, D., Aickin, M., Sennott-Miller, L., Buller, M. K., et al. (1999). Randomized Trial Testing the Effect of Peer Education at Increasing Fruit and Vegetable Intake. *JNCI Journal of the National Cancer Institute, 91*(17), 1491-1500.

Clauset, A. (2005). Finding local community structure in networks. *Physical Review E, 72*(2), 026132.

Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*.

Guimerà, R., Mossa, S., Turtschi, A., & Amaral, L. A. N. (2005). The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences, 102*(22), 7794-7779.

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy2008)*, 11-15.

Harris, K. M. (2008). The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002 [machine-readable data file and documentation]. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.

Kearns, M., Suri, S., & Montfort, N. (2006). An Experimental Study of the Coloring Problem on Human Subject Networks. *Science, 313*(5788), 824-827.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 404-409.

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review, 45*(2), 167-256.

Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), 74*(3), 036104-036119.

Porter, M. A., Onnela, J.-P., & Mucha, P. J. (2009). *Communities in Networks*.

A Measure of Centrality for Dense Networks with Valued Ties

George A. Barnett
Department of Communication, University of California at Davis
Davis, California

This paper presents a new measure of centrality, *scalar products centrality* that is appropriate for dense networks in which link strength is measured with real numbers rather than by a simple dichotomy. Scalar products centrality may be defined by a node's distance from the center of the set of measured relations that compose a network. Formulas for its calculation based upon the centroid scalar products matrix from classical multidimensional scaling are presented. Two examples are provided and the measure is compared with standard measures of centrality (degree, eigenvector centrality, betweenness and closeness) to demonstrate its validity. As expected, the measure is strongly related to the degree and eigenvector measures and less so with betweenness and closeness.

Author
George A. Barnett, Ph.D. (University of California, Davis) is currently Professor & Chair of the Department of Communication at UC Davis, as well as the President of the International Network for Social Network Analysis (INSNA). His research focuses on the structure of international information flows with particular emphasis on the international telecommunications network, how it has evolved as a result of the process of globalization and its role in social, political and cultural change.

Correspondence: *Contact George Barnett at the Department of Communication, University of California, Davis, CA 95616; email; gbarnett@ucdavis.edu*

INTRODUCTION

Almost all network analysis is conducted with sparse networks and dichotomous (binary) ties. This development is partly the result of the reliance on graph theory. Newman (2003, p. 168) defines a network as "…a set of items, which we will call vertices or sometimes nodes, with connections between them, call edges. Systems take the form of networks (also called "graphs" in much of the mathematical literature)." However, network analysis should not be confused with graph theory.

While it is the usual procedure to examine social networks composed of binary graphs, there is no reason that the measurement of a tie should be limited to its presence or absence (Butts, 2009). There are many advantages to this course of action including simplifying the analysis of the network and it's graphic representations. However, the relation between two nodes might be the number of interactions between i and j, or the dollar value of trade from country i to country j, which may be distinct from the value of trade from j to i (Wasserman & Faust, 1994, p. 140-141). Further, Harary (1969) indicates that the lines in "graphs" can take on any positive real number. In fact, perhaps the single most cited article in the area of social networks is Mark Granovetter's "The Strength of Weak Ties" (1973), which implies that ties can take on different values on some attributes.

The standard practice in the analysis of networks is to take valued data and dichotomize the network at the mean or median. That procedure is fraught with the danger of misinterpretation of the network's link structure (Butts, 2009). Cohen (1983) argues dichotomization results in the loss of explained variance and statistical power. Further, "…since methods are available for making use of all the original scaling information, there is no reason to sustain them (the costs in variance accounted for and in power)." (Cohen, 1983, p. 249). Researchers perform dichotomization in order to simplify the data analysis and due to the popularity of loglinear models, which has promoted this

practice among researchers eager to apply state-of-the-art methodologies (Cohen, 1983). In the examples in this paper, a measure is applied to precise behavior data that warrants the use of methods appropriate for continuous variables. Additionally, while there may be measurement error in the data, there is no need to add errors of discreteness (Barnett, Hamlin & Danowski, 1981). After all, dichotomization results in a reduction of the explained variance to 0.647 (Cohen, 1983).

The emphasis on the study of sparse networks is perhaps a result of the widespread finding of the power law degree distribution (Broder, et al., 2000; Barabási, 2002; Barnett & Park, 2005). "...the power law distribution implies that there is an abundance of nodes with only few links, and a small-but significant-minority that have a very large number of links" (Barabási, 2002, p. 71). In other words, most nodes have only a few links, held together by a few highly connected ones that act as network hubs. The later are at the core and the former, the periphery. This is similar to other large networks, such as the Internet.

However, dense networks do exist. For example, one might examine completely connected subgroups rather than entire systems or examine networks of aggregates—composed of the set of ties for a collection of individual nodes that might result from aggregation or from multi-level analysis (Monge & Contractor, 2003; Contractor, Wasserman & Faust, 2006; Vespignani, 2009). Much of my research examines the flow of information among divisions of formal organizations (Barnett & Danowski, 1992; Doelfel & Barnett, 1999), cities (Barnett & Rice, 1985; Choi, Barnett, & Chon, 2006), or nations (Barnett & Park, 2005; Barnett 2001; Salisbury & Barnett, 1999), where each collective entity has ties to all others. Importantly, however, these dense networks do vary in the strength of ties due, in part, to the population of the node.

Because networks of this type have largely been ignored, this paper introduces a measure of

centrality designed particularly for dense networks where the strength of ties may take on any real positive value.

METHODS

1. Application of Existing Measure of Centrality to Dense Networks

While there are numerous measures of centrality, most are inappropriate for dense networks with weighted values. **Betweenness** is the proportion of geodesics linking nodes j and k that pass through node i (Freeman, 1979):

$$C_B(s_i) = \Sigma(P_i(kj)/P(kj))/(n-1(n-2)/2) \qquad (1)$$

where $P_i(kj)$ the number of geodesics between k and j that i lies on, and $P(kj)$ the total number of geodesics between k and j. The betweenness centrality of i is the average across all pairs of nodes (Jackson, 2008).

In dense networks there is little, if any, variance on this measure, since j and k may be connected on paths through most (or even all) other nodes. Further, all geodesics are treated as the same length (Wasserman & Faust, 1994, p.189). Therefore, betweenness does not reveal how nodes differ in centrality for dense networks with weighted values.

The farness of node i is the number of ties that compose its geodesics to all other nodes. The reciprocal of farness is **closeness centrality**:

$$C_c(s_i) = (n-1)/\Sigma l(i,j) \qquad (2)$$

where $l(i,j)$ is the number of links in the shortest path between i and j (Jackson, 2008). Since closeness does not consider the weighted values of links, in dense networks all nodes are equidistant from all other nodes. Thus, this measure is inappropriate for dense networks with weighted values.

Two different measures of centrality have been commonly used with dense networks where s_{ij} may be any real number. They are degree centrality (Freeman, 1979), and eigenvector

centrality (Bonacich, 1972). **Degree** is simply the sum of the values of each row (or column) (Borgatti, 2005):

$$C_D(s_i) = \Sigma s_{ij} \qquad (3)$$

The disadvantage of degree is that the only information it provides is the sum of the weights. It says nothing about the distribution of the network's link strengths. A node could have a very strong relationship with only one node and thus be very central, while residing at the periphery of the network.

According to Wasserman and Faust (1994, pp. 193-197) **information centrality** is the generalization of betweenness centrality such that all paths between the nodes weighted by their length (link strength) are considered when calculating the measure. It is calculated using Formula 4:

$$C_I(n_i) = 1/(c_{ii} + (T - 2R)/g) \qquad (4)$$

Where $T = \Sigma\ c_{ii,}$ the sum of the diagonal elements or the trace of the socio-matrix, and $R = \Sigma\ c_{ij,}$ the sum of any one of the rows. Stephenson and Zelen (1989) recommend using relative information indices, which may be obtained by dividing $C_I(n_i)$ by the total of all the indices:

$$C'_I(n_i) = C_I(n_i)/\Sigma_i C_I(n_i) \qquad (5)$$

Like betweenness, information centrality is problematic. In dense networks there is little, if any, variance on this measure, since j and k may be connected on paths through most (or even all) other nodes, although this measure does consider the sum of the link strengths in its calculation.

Eigenvector centrality is defined as the largest eigenvector of the socio-matrix defining the network (Bonacich, 2007). The defining equation of the eigenvector is:

$$\lambda\mathbf{v} = \mathbf{Sv} \qquad (6)$$

where $\mathbf{S}$ is the socio-matrix of the network, λ is the largest eigenvalue of $\mathbf{S}$ and $\mathbf{v}$ is the

eigenvector. Bonacich's measure is accurate only to the extent that the largest eigenvalue accounts for a large proportion of variance in the network. In dense networks, the largest eigenvalue may account for as little as $1/(n-1)*(100)$ per cent of the variance in the socio-matrix. In a relatively small 25 node network, this may be as little as 4.17%.

In the section that follows, an alternative measure of centrality, Scalar Products Centrality will be presented. It has none of the disadvantages of degree or eigenvalue centrality for the examination of dense networks with link strength having values other than zero and one.

2. Scalar Products Centrality

Scalar products centrality may be defined by a node's distance from the center of the set of measured relations that compose a network. Its calculation is based upon the centroid scalar products matrix, the first step in classical multidimensional scaling (Torgerson, 1958). Its computation assumes that **S** is a one-mode square socio-matrix of social distances, such as the frequency of communication reversely scaled (Barnett, 1988). The diagonal, $s_{ii} = 0$, since the distance of a node from itself is by definition zero. The first step is:

$$\mathbf{B} = \mathbf{SS^T} \qquad (7)$$

where, **B** is a matrix of squared distances. However, the origin of **B** is at point i. **B** must be translated to the centroid of all points. This is accomplished by "double centering" **B,** i.e.**,** subtracting the row and column means of the matrix from its elements, adding the grand mean and taking the square root of **B**. The centered **S*** may be defined by subtracting the row and column means from the elements of **S.** Thus, the scalar products of the centered configuration is:

$$\mathbf{B^*} = \mathbf{S^*S^*} \qquad (8)$$
or
$$\mathbf{B^*} = (\mathbf{B} - \mathbf{B_{+j}} - \mathbf{B_{+j}}^T + \mathbf{B_{gm\mu}}) \qquad (9)$$

where, $\mathbf{B_{cm+j}}$ is a matrix of row or column means of **B**, and $\mathbf{B_\mu}$ contains the grand mean of all cells of **B**.

This assumes that **S*** is a symmetrical (non-directional) matrix. The scalar version of equation 7 defines each individual element of **B*** as (Torgerson, 1958, p. 258):

$$b_{ij}^* = \tfrac{1}{2}(1/k(\sum_{i}^{k} s^2_{ij}) + 1/k(\sum_{j}^{k} s^2_{ij}) - (1/k^2)\sum_{i}^{k}\sum_{j}^{k} s^2_{ij} - s^2_{ij})$$

$$(10)$$

The centralities of the individual nodes are the absolute values of the square root of the diagonal, $\left| \sqrt{b_{ii}} \right|$, their distances from the origin of the distribution. These values may be normalized by dividing by the largest element such that the least central node is equal to one.

One advantage of the scalar products measure is that it is isomorphic with the theoretical concept of centrality. In a multidimensional space, centrality may be defined as the distance from the center of the space. Further, **B***, the "scalar products" is the first step in the calculations for classical multidimensional scaling (Torgerson, 1958) and $\left| \sqrt{b_{ii}} \right|$ is the distance of the individual node from the origin of the distribution.

This measure has been used in the past by Barnett and Rice (1985) to determine the relative centralities of cities within the domestic air traffic network and a computer–based conference group. Barnett and Danowski (1992) applied the measure to examine the structure of the field of Communication based on scholars' affiliations in a professional organization.

3. Examples

3.1 Canadian Interprovincial Migration

To demonstrate the utility of the scalar products measure, inter-provincial migration data for January to December 1998 from Statistics Canada (Bélanger, 1999) were employed. The

Table 1. Socio-matrix of Canada's Interprovincial Migration

		1	2	3	4	5	6	7	8	9	10	11	12	13
1	**Newfoundland**	0	42	570	242	42	1112	57	59	906	187	0	18	41
2	**P.E.I.**	42	0	207	122	42	272	25	15	171	124	0	0	7
3	**Nova Scotia**	570	207	0	932	261	2397	278	95	919	592	42	43	84
4	**New Brunswick**	242	122	932	0	573	2197	278	95	919	592	42	13	16
5	**Quebec**	42	42	261	573	0	5512	231	124	915	826	25	0	59
6	**Ontario**	1112	272	2397	2197	5512	0	1517	617	4244	4187	82	140	25
7	**Manitoba**	57	25	278	278	231	1517	0	760	1670	1125	9	37	22
8	**Saskatchewan**	59	15	95	95	124	617	760	0	2999	860	23	47	13
9	**Alberta**	906	171	919	919	915	4244	1670	2999	0	6617	97	340	42
10	**B.C.**	187	124	592	592	826	4187	1125	860	6617	0	190	127	53
11	**Yukon**	0	0	42	42	25	82	9	23	97	190	0	50	6
12	**NWT**	18	0	43	13	0	140	37	47	340	127	50	0	44
13	**Nunavut**	41	7	84	16	59	25	22	13	42	53	6	44	0

Table 1. Data was symmetrized by taking the means of s_{ij} and s_{ji}.

data on the number of migrants are based on Revenue Canada Tax and child tax credit files. This is a relatively dense (though not completely interconnected) network (density = .949), with people moving among all provinces or territories except between the Yukon and Newfoundland and Prince Edward Island, and between the Northwest Territories and Prince Edward Island and Quebec. The greatest flow of migrants was between Alberta and British Columbia, 6,617 people. It was followed by Ontario and Quebec, 5,512. Table 1 presents the socio-matrix of this network. Figure 1 provides a graphic representation of the network.

Table 2 presents the scalar products centralities for the Canadian provinces and territories. These were calculated from the socio-matrix (Table 1) converted first to social distances by reverse scaling. The diagonal, s_{ii}, remained as zeros. Included in the table are the nodes' squared distances from the origin, the distances

and the normalized values. The normalized values were calculated by dividing the distances by the distance for Nunavut - the most peripheral node. Further, one can convert this measure of distance (eccentricity) to a measure of centrality in which the most central nodes has the greatest value by simply subtracting by 1.0. Alternatively, one may normalize the central scalar products measure by taking the inverse of the square root of the distance and multiplying by a scaling constant (k). For the example in Table 2, the values were multiplied by 1,000.

As can be seen in Table 2, Ontario is the most central node, followed by Alberta and British Columbia. At the periphery of the network are the three territories, Yukon, Northwest Territories and Nunavut.

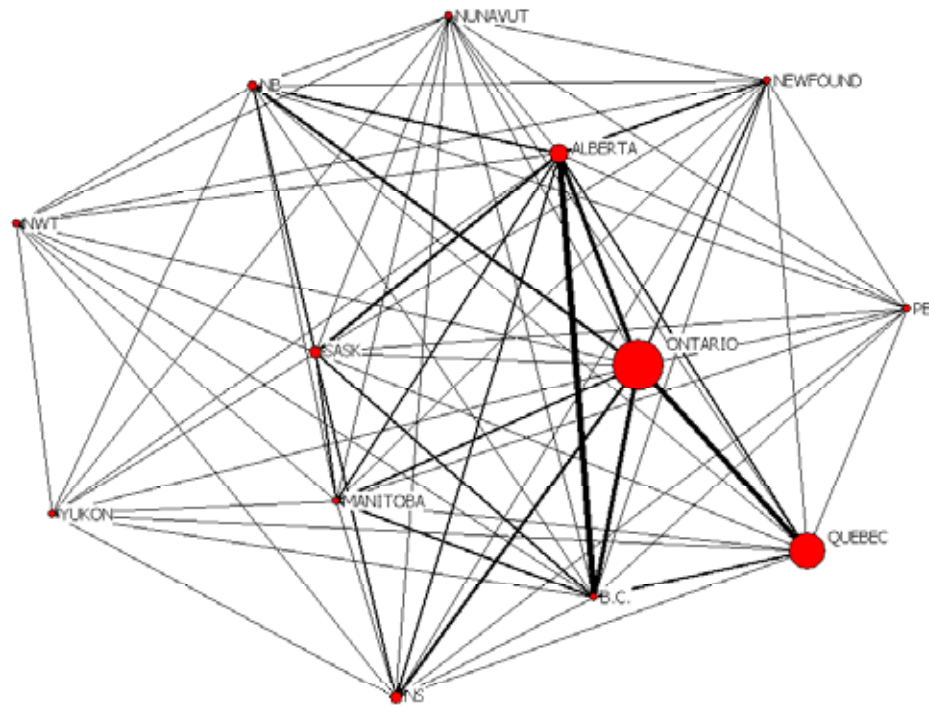**Figure 1. Canada's Inter-provincial Migration Network**

Figure 1. The thickness of the link indicates the amount of migration between provinces. The diameter of the nodes indicates their relative population.

Table 2. Scalar products Centrality for Canadian Provinces

Province/ Territory	Distance b_{ii}	Sqrt. $b_{ii}^{1/2}$	Normalized $b_{ii}^{1/2}$/max	Centrality $1-b_{ii}^{1/2}$/max	Inverse $k(1/b_{ii}^{1/2})$
Newfoundland	19982360	4470.16	0.9380	0.0620	0.2237
P.E.I.	22094350	4700.46	0.9863	0.0137	0.2127
Nova Scotia	17232080	4151.15	0.8711	0.1289	0.2409
New Brunswick	17564310	4190.98	0.8794	0.1209	0.2386
Quebec	16851300	4105.03	0.8614	0.1386	0.2436
Ontario	6616079	2572.17	0.5397	0.4603	0.3888
Manitoba	17558550	4190.29	0.8793	0.1207	0.2386
Saskatchewan	18141740	4259.31	0.8938	0.1067	0.2348
Alberta	8859242	2976.45	0.6246	0.3754	0.3360
B.C.	12352640	3514.63	0.7375	0.2625	0.2845
Yukon	22554330	4749.14	0.9966	0.0034	0.2106
NWT	22263910	4718.46	0.9901	0.0099	0.2119
Nunavut	22708080	4765.30	1.0000	0.0000	0.2099

Table 3. Correlations of Scalar Products Centrality with Other Measures

Measure	Distance b_{ii}	Sqrt. $b_{ii}^{\frac{1}{2}}$	Normalized $b_{ii}^{\frac{1}{2}}$ /max	Centrality $1- b_{ii}^{\frac{1}{2}}$ /max	Inverse $k(1/b_{ii}^{\frac{1}{2}})$
Degree	-0.995	-0.994	-0.994	0.994	0.994
Eigenvector	-0.977	-0.964	-0.964	0.964	0.964
Closeness	0.598	0.565	0.565	0.565	0.565
Betweenness	-0.585	-0.556	-0.556	-0.556	-0.556
Information	-0.985	-0.968	-0.968	-0.968	-0.968

Table 4. Correlations of Scalar Products Centrality with other Measures

Measure	Distance b_{ii}	Sqrt. $b_{ii}^{\frac{1}{2}}$	Normalized $b_{ii}^{\frac{1}{2}}$ /max	Centrality $1- b_{ii}^{\frac{1}{2}}$ /max	Inverse $k(1/b_{ii}^{\frac{1}{2}})$
Degree	-0.948	-0.895	-0.895	0.895	0.895
Eigenvector	-0.937	-0.891	-0.891	0.891	0.891
Information	-0.944	-0.940	-0.940	0.940	0.940
Closeness	constant				
Betweenness	constant				

Correlations with other measures of centrality.
Table 3 provides the correlations of scalar products centrality with the other measures of centrality.

Worth noting are scalar products centrality's nearly equivalent relationships with degree and eigenvector measures, the two measures generally used to describe dense networks with measured link strengths. This is because the more central nodes, Ontario, Alberta and British Columbia had the strongest ties overall (greatest migration) and the peripheral nodes, the weakest (least migration). The first eigenvalue from which the eigenvector measure was calculated accounted for only 19.5% of the variance in the network and the second 14.3%[1], suggesting that the Bonacich measure is considering only part of
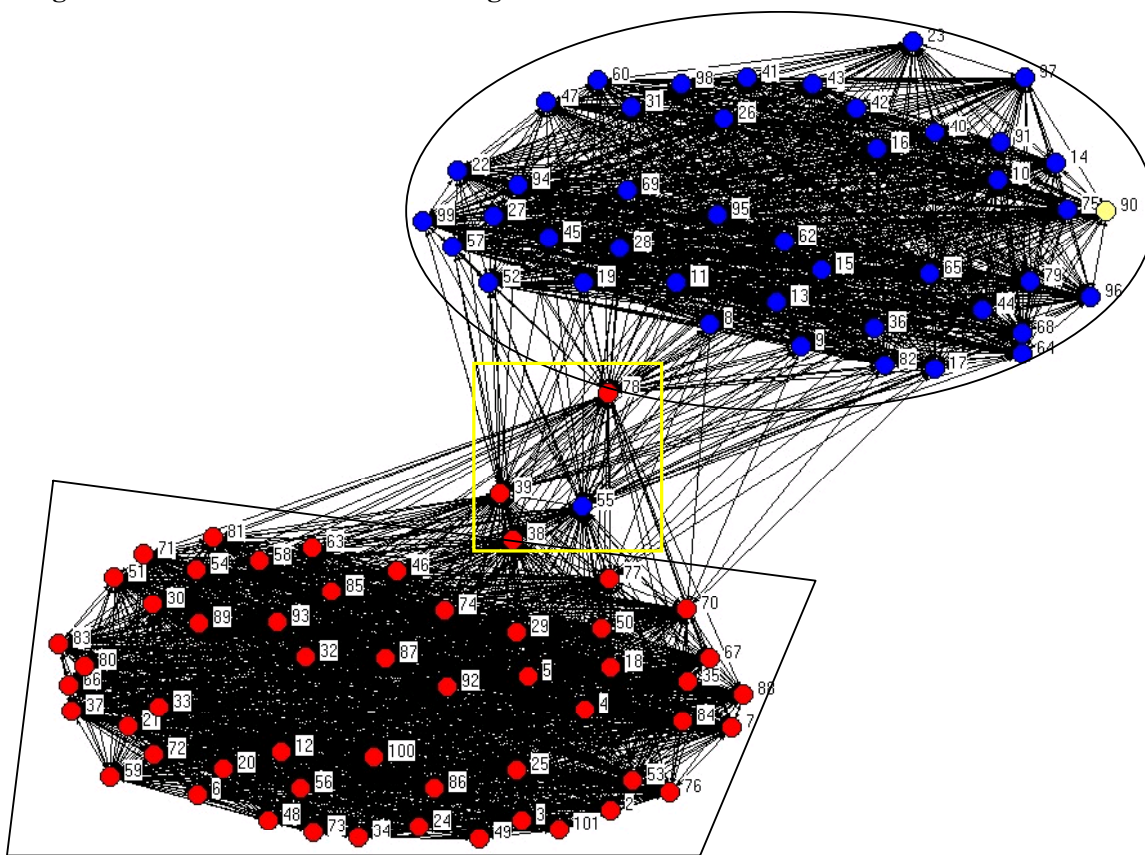
the relations in the network. Note also that the correlation with closeness and betweenness are much lower. This is because the nodes vary little in closeness. All but four ties are one-step distant. The other four have two-step links. Likewise, there is very little variance in betweenness. There are many pathways around any individual node due to the network's density.

These five relationships should be taken as an indicator of the validity of the scalar products measure. It is strongly related to those measures which it theoretically should be and not strongly related to those which it shouldn't.

Similar patterns were found by Valente, et al. (2008), who reported that for sparse networks degree and eigenvector centrality were strongly correlated with a weaker relationship with betweenness and closeness. Since scalar products centrality is strongly related to

---

[1] This is only the variance on those dimensions with positive (real) eigenvalues. See Barnett and Rice (1985) for a further explanation.

Figure 2. United States Senate Voting Network

eigenvector centrality both theoretically and empirically, this pattern of relations would be expected.

3.2 United States Senate Voting

The Canadian migration network was composed of a single group. Would the scalar products measure of centrality prove equally valid as a measure of centrality for dense networks with a more complex structure? Figure 2 graphically presents the voting pattern of the 109th U.S. Senate. Roll call voting data were acquired from the United States Senate, *Legislation and Records* Website (*http://www.senate.gov/legislative/LIS/roll_call_lists*). This is the United States Senate's official source.

Although all 100 senators voted in common over 200 times (density = 1.000), the network is strongly clustered by party affiliation. At the top is the cluster of the Democratic caucus. At the bottom are the Republicans. In the center is a group composed of less partisan senators. There are four nodes in the middle of the two groups: Susan Collins (R, ME, #38), Olympia Snowe (R, ME, #39), Ben Nelson (D, NE, #55) and Lincoln Chafee (R, RI, #78).

Table 4 provides the correlations of scalar products centrality with the other measures of centrality for the U.S. Senate network. The correlations with the closeness and betweenness measures are undefined because these indicators are constant for completely interconnected networks. This clearly shows that these measures of centrality are inappropriate for dense networks. The first eigenvalue accounted

for 24.8% of the variance (the second, 6.4%) in the U.S. Senate network [2]. It has a weaker relation, .895 as compared to .964, with the scalar products measure. This indicates the uniqueness of the scalar products indicator for completely dense networks with complex structures composed of multiple groups.

CONCLUSIONS

This paper presented a new measure of centrality, scalar products centrality that is appropriate for dense networks in which link strength is measured. Scalar products centrality may be defined by a node's distance from the center of the set of measured relations that compose a network. The formulas for its calculation (based upon the centroid scalar products matrix of classical multidimensional scaling) were presented. Examples were provided and the measure was compared with standard measures of centrality (degree, eigenvector centrality, betweenness and closeness) to demonstrate its validity.

As expected, the method is strongly related to the degree and eigenvector measures, the standard measures for valued data, and less so to betweenness and closeness. Further, the more complex the network structure, the greater the difference between scalar products centrality and the eigenvector and degree centrality. It is in situations where networks have a complex structure with many tightly connected groups interconnected by relatively weaker ties that scalar products centrality is most appropriate.

REFERENCES

Barabási, A. (2002). *Linked: The new science of networks*. Cambridge, MA: Perseus Publishing.

Barnett, G.A. (1988). Approaches to non-Euclidean network analysis. In *Small World*. M. Kochen (Ed.), Norwood, NJ: Ablex, 349-372.

Barnett, G. A. (2001). A longitudinal analysis of the international telecommunication network, 1978–1996, *American Behavioral Scientist*, 44, 1638–55.

Barnett, G.A. & Danowski, J.A. (1992). The structure of communication: A network analysis of the International Communication Association. *Human Communication Research*, 19, 264-285.

Barnett, G.A., Hamlin, D.M., & Danowski, J.A. (1981). The use of fractionation scales in communication audits. In M. Burgoon (Ed.), *Communication Yearbook 5* (pp. 455-471), New Brunswick, NJ: Transaction.

Barnett. G.A., & Park, H.W. (2005). The structure of international internet hyperlinks and bilateral bandwidth. *Annals of Telecommunication*, 60, 1115-1132.

Barnett, G.A. & Rice, R.E. (1985). Network Analysis in Riemann Space: Applications of the Galileo System to Social Networks. *Social Networks*, 7, 287-322, 1985.

Bélanger, A. (1999). Report on the demographic situation in Canada 1998-1999: Current demographic analysis. Statistics Canada (Catalogue no. 91-209-XPE), Ottawa.

Bonacich, P. (1972). Techniques for analyzing overlapping memberships. *Sociological Methodology*, 4, 176-185.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, 29, 555-564.

Borgatti, S.P. (2005). Centrality and network flow. *Social Networks*, 27, 55-71.

Broder, A., Kumar, R, Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., & Wiener, J. (2000). Graph structure on the web. *Computer Networks*, 33, 309-320.

Butts, C.T. (2009). Revisiting the Foundations of Network Analysis, *Science*, 24(325), 414-416.

Choi, J.H., Barnett, G.A., & Chon, B.S.(2006). Comparing world city networks: A network analysis of Internet backbone and air transport intercity linkages, *Global Networks*, 6(1), 81-99.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7(3), 249-253.

---

[2] Again, this is only the variance on the positive eigenvalues

Contractor, N.S., Wasserman, S. & Faust, K. (2006). Testing multi-level, multi-theoretical hypotheses about networks in 21st century organiztional forms: An analytic framework and empirical example. *Academy of Management Review, 31*, 681-703.

Doerfel, M.L. & Barnett, G.A. (1999). A comparison of the semantic and affiliation networks of the International Communication Association. *Human Communication Research, 25* (4), 589-603.

Freeman, L. C. (1979). Centrality in Social Networks: Conceptual Clarification. *Social Networks* 1, 215-239.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 81*, 1287-1303.

Harary. F. (1969). *Graph Theory*. Reading, MA: Addison-Wesley.

Jackson, M.O. (2008). *Social and economic networks*. Princeton, NJ: Princeton University Press.

Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. New York: Oxford University Press.

Newman, M.E.J. (2003). The structure and function of complex networks. *SIAM Review, 45(2),* 167-256.

Salisbury J. & Barnett, G. A. (1999). The World System of International Monetary Flows: A Network Analysis, *The Information Society*, 15, 31-49.

Torgerson, W.S. (1958). *Theories and methods of scaling*. New York: Wiley.

Valente, T.W., Coronges, K., Lakon, C. & Costenbader, E. (2008). How correlated are network centrality measures? *Connections, 28(*1). 16-26.

Vespignani, A. (2009). Predicting the Behavior of Techno-Social Systems. *Science* 24(325), 425-428.

Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications*, New York: Cambridge University Press.

Network Topology Effects on Correlation between Centrality Measures

Ian McCulloh

United States Military Academy,
Network Science Center, West Point, NY

Centrality measures are often used to describe influential nodes in a network. When these measures are highly correlated they may be redundant and when they are uncorrelated they provide unique insight into the network. I propose a network simulation approach that creates networks with varying degrees of Erdos-Renyi randomness and Albert-Barabasi scale-freeness. Using this simulation approach I conduct 10 replications of a full factorial experimental design with varying levels of density and randomness versus scale-freeness. The effects of topology and density on the correlations of degree, betweenness, closeness, and eigenvector centrality are investigated. I find that not only does density and topology affect the correlation of centrality measures, but there exist many interaction effects as well. In general, networks with high Erdos-Renyi randomness tend to have higher levels of correlation between centrality measures than networks with Albert-Barabasi scale-freeness.

Author
Ian McCulloh, Ph.D. is an Assistant professor in the Department of Behavioral Sciences and Leadership at the United States Military Academy.

Correspondence: *Contact Ian McCulloh, Department of Behavioral Sciences and Leadership, at the United States Military Academy, 646 Swift Road, West Point, NY 10996; email: ian.mcculloh@usma.edu.*

INTRODUCTION

Centrality measures are often used in social network analysis to identify influential nodes. Among the most common centrality measures are degree, closeness, betweenness (Freeman, 1979), and eigenvector (Bonacich, 1972) centralities. Each of these centrality measures is based on a different concept of what makes a node influential or central to a network. When these measures are highly correlated, the differences in centrality measures are insignificant and redundant (Valente et al, 2008). When there is low correlation between centrality measures, they may offer unique insight into the network.

There have been a few studies that investigate the correlation between network centrality measures. Three studies investigated the correlation of measures under conditions of missing data (Bolland, 1998; Borgatti et al 2006; McCulloh, 2009). Other studies investigated the correlation between network centrality measures in applied network studies (Rothenberg, 1995; Faust, 1997; Valente and Forman, 1998). A more recent study investigated the correlation of centrality measures in empirical network data (Valente et al, 2008). This was an important study for revealing the effects of correlation on social network analysis. Valente also identifies the affect of density and directionality on the correlation of centrality measures. This paper expands upon their research by exploring how the underlying network topology effects these correlations. In using the concept of network topology, the findings can be more easily generalized to new network studies.

The idea of network topology is not clearly defined in the literature. The concept is used here only in the context of the degree distribution. Two network topologies are therefore compared: random and scale-free networks. Paul Erdős and Alfréd Rényi (1959) proposed the random network. In this model nodes are connected with some probability, p. Others have shown that the distribution of the degree will follow a binomial distribution (Bolobos, 2003, McCulloh, 2009). Albert-Lazlo Barabasi and Reka Albert (1999) proposed an alternate model of network topology based on the theory of preferential attachment. The mechanism of preferential attachment was originally proposed by Herb Simon (1955) and suggests that when social groups are growing, new members will choose to connect to individuals with probability proportional to their prestige, often measured by degree (Wasserman and Faust, 1994). Simon, and later Barabasi and Albert, show that when preferential attachment exists in a growing network, the distribution of the degree will follow a power law. Under certain conditions this creates a network with a few central hub nodes. It has been shown that the effect of missing data on network measures is significantly affected by the degree topology of the network (Frantz et al, 2009). Networks that have strong tendencies for preferential attachment are more robust in their identification of top central nodes when there exists increasing levels of missing data.

In this paper, a simulation methodology is proposed that will generate networks that fall on a continuous spectrum between Erdős-Rényi random and Albert-Barabasi scale-free. Using this simulation methodology, networks are generated where the topology and density of the network are varied in a full-factorial, statistically designed experiment to explore the response surfaces of the correlation between network centrality measures.

In the next section the four centrality measures are briefly reviewed: degree, closeness, betweenness and eigenvector centrality. These measures were chosen for this study because they are the most common in the literature. The Methods section provides a description of how networks are simulated and an explanation of the virtual experiment. The results will be presented followed by discussion.

BACKGROUND

The *degree centrality* of node a is the number of other nodes directly connected to node a. The

degree centrality is a measure of direct influence in the network.

The *closeness centrality* of a node describes how close a node is to all others in the network. To calculate closeness, the geodesic path between all pairs of nodes in the network must be calculated. The closeness of node *a* then is the inverse of the average geodesic from node *a* to all other nodes in the network. A node high in closeness can disseminate information within the network more efficiently based on its position. It is therefore an ideal target for network diffusion.

The *betweenness centrality* of a node describes how frequently a node falls on the geodesic between other nodes. The betweenness of node *a* is the number of geodesic paths containing node *a* dived by the total number of geodesics in the network which is always $n(n-1)$, where *n* is the number of nodes in the network and we assume that there are no reflexive links. A node that is high in betweenness is influential in that it can connect otherwise disconnected subgroups within the network. High betweenness nodes serve as gate keepers or brokers of information and resources.

Eigenvector centrality is based on the concept that a node is influential to the extent that it connects to influential alters. Bonacich (1972) discovered that if the degree centrality of a node is modified by the degree centrality of its alters and this process is iterated many times, the result converges to the eigenvector of the symmetric adjacency matrix of the network.

The density of a network is the number of links in the network divided by the number of possible links which is always $n(n-1)$ when we assume that there are no reflexive links in a directed network. It can be shown that the maximum likelihood estimate of the parameter *p* in an Erdős-Rényi random network is the density. This relationship allows the density to be specified as an experimental factor in the statistically designed experiment to explore the density and degree topology effects on centrality measure correlation.

METHODS

Networks can be simulated in a manner that allows them to achieve a hybrid topology between Erdős-Rényi randomness and Albert-Barabasi scale-freeness. Here, this is achieved by randomly assigning links between nodes as in a random graph for some percentage of the target order of the network. Then, the remaining nodes are added using preferential attachment. In this study, all networks consist of 100 nodes. For example, in the first virtual experiment, 20 percent of the nodes, or 20 nodes are randomly assigned links with a probability, $p = 0.1$ corresponding to the target density. The remaining 80 nodes are added to the network using preferential attachment. This results in a network that is 20 percent random and 80 percent scale-free. In the second virtual experiment 80 nodes are assigned random links, while 20 are added via preferential attachment. This results in a network that is 80 percent random and 20 percent scale-free. In this manner, networks can be simulated that have varying levels of randomness and density. These network variables are then varied in a statistically designed experiment.

A two factor, full factorial, statistically designed experiment is used to explore the response surface of the correlations between centrality measures as a function of the topology and density. The response variables are the 6 correlations between the four centrality measures. The independent factors are the network topology and the density. The high and low level of the topology is set at 80 percent and 20 percent random, respectively. The high and low level of the density is set at 0.3 and 0.1 respectively. A center point was also used, where the topology is set at 50 percent random and the density is 0.2, in order to estimate significant curvature in the response surface. 10 replications at each combination of independent factors are run to estimate standard error and make accurate inference. Table 1 displays the

design matrix for the series of virtual experiments. "Std Ord" refers to the standard order of the full factorial experimental design (Montgomery, 2005).

Table 1. Experimental Design

Std Ord	Reps	No. Nodes	Initially Random Nodes	Target Density
1	10	100	20	0.1
2	10	100	80	0.1
3	10	100	20	0.3
4	10	100	80	0.3
5	10	100	50	0.2

RESULTS

Topology and randomness/scale-freeness are all highly, statistically significant predictors of correlation between centrality measures for five of the six correlations. The randomness/scale-freeness factor does not appear to be significant in the degree centrality – eigenvector centrality correlation, however, there does appear to be an interaction effect. The half normal plot for the degree centrality – betweenness centrality correlation is presented in Figure 1.

A half normal plot is used to demonstrate the significance of independent factors on the correlation value. The blue squares represent the independent factors and the green triangles represent the degrees of freedom associated with the error term. The factors represented in the half normal plots are all highly significant with p-values less than 0.001. The notation "A" is used in the plot to represent the density factor. The notation "B" is used to represent the randomness/scale-freeness factor. The notation "AB" represents the interaction effect between the two other factors. An interaction implies that one factor has a different effect on the response, depending on the level of the other factor.

The half normal plots for the other correlations between centrality measures are similar to the

Figure 1. HNP Correlation Degree-Betweenness
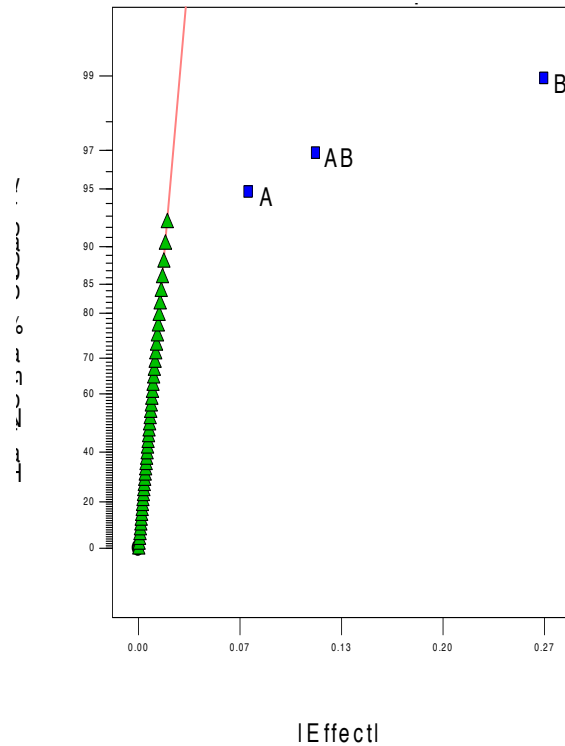
|Effect|

Figure 1. HNP (Half Normal Plot) is a graphical tool that uses ordered estimated effects to help assess which factors are significant and which are not.

degree – betweenness correlation; however, there are no interaction terms in the correlations involving closeness centrality. In addition, both independent factors (density and randomness) are significant in all correlations except for the degree-eigenvector centrality. For this correlation, the topology is only present in the interaction term.

It is important to also consider the degree to which topology and density affect the correlation between measures. Table 2 displays the coefficient of determination for the six response surfaces corresponding to the six correlations between centrality measures. The coefficient of determination quantifies how much of the variability in the correlation can be explained by the factor's topology, density, and

interaction. It can be seen in Table 2 that the coefficients of determination are all very high, indicating that topology and density are perhaps the most significant factors affecting the correlation between centrality measures.

Table 2. Coefficients of Determination

Correlation	Coefficient of Determination
Degree-Betweeness	0.9474
Degree-Closeness	0.9178
Degree-Eigenvector	0.7812
Betweeness-Closeness	0.8526
Betweeness- Eigenvector	0.9221
Closeness-Eigenvector	0.9131

The actual response surfaces for the correlation between centrality measures are presented in Table 3. These equations can be used to calculate the expected correlation between centrality measures in a 100 node network, given a density and Erdos-Renyi randomness. The sign and magnitude of the coefficients in the

equation also provide insight into the effects of density and topology on the correlations. For example, the intercept term in the Degree-Eigenvector correlation is 0.95, which is an extremely high correlation. The coefficients for randomness and the interaction effect are small. This suggests that the density of the network has the largest effect on correlation between these measures. Because the coefficient of the density term is positive, the correlation between these measures will increase as the density increases.

A counter example of the Degree-Eigenvector correlation is found in the Degree-Closeness correlation. The intercept term for this response surface is low at 0.20. The coefficient for the randomness term is relatively high and negatively signed. This means that high randomness will lead to a more negative correlation between degree and closeness. In addition, an increase in the density of the network will contribute to negative correlations between the degree and closeness centrality measures.

Table 3. Response Surface Equations for the Correlation between Centrality Measures

Correlation	Final Equation in Terms of Actual Factors
Degree-Betweeness	$= 0.87 - 1.35d + 0.056r + 1.97dr$
Degree-Closeness	$= 0.21 - 0.55d - 0.74r$
Degree-Eigenvector	$= 0.95 + 0.12d + 0.01r - 0.08dr$
Betweeness-Closeness	$= -0.09 - 0.44d - 0.43r$
Betweeness- Eigenvector	$= 0.77 - 1.07d + 0.11r + 1.72dr$
Closeness-Eigenvector	$= 0.28 - 0.64d + 0.79r$

* d is the density of the network, and r is the Erdos-Renyi randomness of the network.

The relationships between the density, the randomness and the interaction effects can be illustrated with an interaction plot found in Figure 2. Figure 2 shows the interaction plot for the Degree – Betweenness centrality interaction. This figure plots the correlation of the two respective measures along the y-axis and the

randomness along the x-axis. There are two lines drawn in the plot corresponding to the high and low levels of network density. The points aligned with the middle of the x-axis are the center points in the experimental design. A significant interaction effect will cause the two lines to have different slopes, showing a different effect dependent upon the other factor.

There is a similar interaction for all of the other correlations with the exception of those involving closeness centrality. Since there is no interaction effect for closeness centrality, the two lines representing the different levels of density are parallel.

The correlations involving betweenness centrality reveal that increased density magnifies the effects of network topology on the correlation. As the density of the network and the randomness of the network increases, the betweenness centrality measure becomes more correlated with degree and eigenvector centrality. Recall that degree and eigenvector centrality are highly correlated for all networks in the study, so a measure correlated with degree will also be correlated with eigenvector centrality.

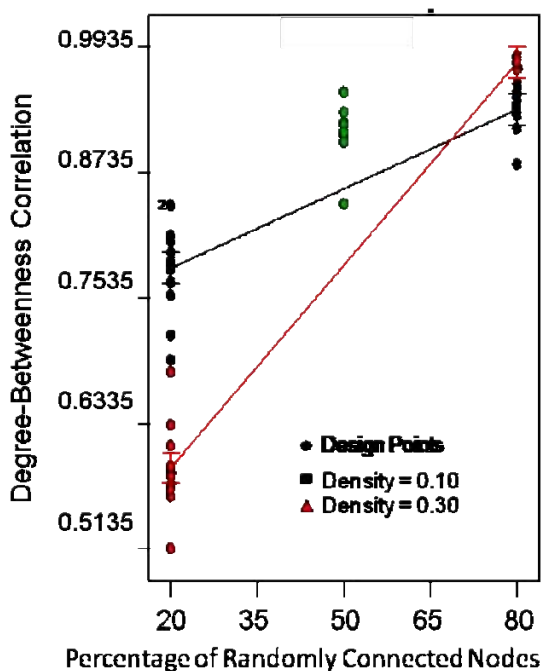**Figure2. Interaction Plot Degree-Betweenness**

Figure 2. The X-axis is the percentage of nodes connected as an Erdős-Rényi Random graph (not preferential attachment). The Y-axis is the correlation between degree and betweenness centrality. The two lines correspond to the density. The line with greater slope has larger density, indicating that as density increases, topology has a greater affect on the correlation of centrality measures.

CONCLUSIONS

The nature of centrality measures in social network analysis is an understudied topic. The few papers investigating this area suggest that an improved understanding of the correlations between network measures are necessary to identify redundant measures versus measures that provide unique insight, as well as to understand the robustness of measures to missing data.

This paper shows that both network density and topology affect the correlation between centrality measures in the network. The correlations between measures of degree, betweenness, and eigenvector centrality exhibit interaction affects between the density and topology. Most compelling is that these two variables explain a vast majority of the variability in the correlation between measures. Finally, simulation provides a controlled environment for virtual experiments to explore network correlation studies, including a novel approach for modeling the scale-freeness of a network.

Only a limited number of measures were investigated in this paper. Future studies may include a wider range of network measures. There are many social network analysis software packages that include these network measures. The Organizational Risk Analyzer (ORA) available from Carnegie Mellon University provides most of these measures as well as a feature to generate stylized networks with varying levels of randomness or scale-freeness. In addition, this work only investigates networks with 100 nodes. It is possible that the order of the network will also affect the correlation of these measures. Based on the relationship between the order of the network and the density, it is unlikely that this would form a linear response surface. Hopefully, future researchers will consider other factors that might contribute significantly to the correlation between network measures.

The effect of topology on centrality measure correlation highlights its importance as a network property. Future investigations of network robustness to missing nodes, links, or attributes should consider network topology. In addition, simulation studies are likely to provide the most fruitful insight into network robustness in a similar approach as used by Borgatti and colleagues (2006). Their study of network robustness did not consider the topology of the network and was focused on Erdos-Renyi random networks. This study demonstrates the need for their important research to be replicated considering the affects of topology on the network.

In light of much discussion at academic conferences in regards to scale-free networks and preferential attachment, it is important to point out that this paper does not address how "real-world" networks evolve or in what ways networks exhibit randomness or scale-freeness. Many have commented on the issue of network topology in the literature and at academic conferences (Alderson, 2008; Doyle et al, 2005; Barabasi, 2008). The issue brought to light in this research is that network topology considerations will significantly affect the correlation between centrality measures. They are also likely to affect the outcomes of robustness studies. Therefore, future research should take network topology into consideration when presenting findings.

The correlation between network measures appears to be more correlated in random networks than in scale-free networks where hub nodes exist. This affect seems to be magnified by the density of the network when network density is high. This suggests that network measures may be redundant for Erdos-Renyi random networks, but that in other networks they may provide valuable, unique insight. Future research involving simulated networks should consider the impact of topology on their findings.

REFERENCES

Alderson, D.L. (2008). Catching the "Network Science" Bug: Insight and Opportunity for the Operations Researcher. *Operations Research*, 56(5), 1047-1065

Barabasi, A.L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509-512.

Barabasi, A.L. (2008). Keynote presentation at the 3rd US Military Academy Network Science Workshop, West Point, NY, 15 October 2008.

Bolland, J.M. (1998). Sorting Out Centrality: An analysis of the Performance of Four Centrality Models in real and simulated networks, *Social Networks*, 10, 233-253.

Bollobas, B. (1998). *Modern Graph Theory*. Springer-Verlag, New York.

Bonacich, P. (1972). Technique for Analyzing Overlapping Memberships, *Sociological Methodology*, 4, 176-185.

Borgatti, S.P., Carley, K.M., & Krackhardt, D. (2006). On the Robustness of Centrality Measures Under Conditions of Imperfect Data, *Social Networks*, 28, 124-136.

Carley, K. M. (2009). *ORA: Organizational Risk Analyzer v.1.9.5.* [Network Analysis Software]. Pittsburgh: Carnegie Mellon University.

Doyle, J., Alderson, D., Li, L., Low, S., Roughan, M., Shalunov, S., Tanaka, R., & Willinger. W. (2005). The "robust yet fragile" nature of the Internet. In: *Proceedings of the National Academies of Science*, 102(41), 14497–14502.

Erdos, P, & Renyi, A. (1959). On random graphs. *Publicationes Mathematicae Debrecen*, 6, 290-297.

Faust, K. (1997). Centrality in affiliation networks. *Social Networks*,19, 157-191.

Frantz, T., McCulloh, I., & Carley, K.M. (2009). Estimating the Reliability of Top-Actor Identification. *Network Science*, 2, to appear.

Freeman, L. (1979). Centrality in social networks: I, conceptual clarification. *Social Networks*, 1 (1979), 215-239.

Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry,* 40 (1977), 35-41.

McCulloh, I. (2009). *Detecting Changes in a Dynamic Social Network*. Ph.D. Dissertation, Carnegie Mellon University.

Montgomery, D.C. (2005). *Design and Analysis of Experiments*. 5th Ed. Wiley, Hoboken, NJ.

Rothenberg, R. Potterat, J.J., Woodhouse, D.E., Darrow, W.W., Muth, S.Q., & Klovdahl, A.S. (1995). Choosing a centrality measure: Epidemiologic correlates in the Colorado Springs study of social networks. *Social Networks*, 17,273-297.

Simon, H. (1955). On a class of skew distribution functions. *Biometrika* 42, 435–440.

Valente, T.W., & Foreman, R.K. (1998). Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. *Social Networks*, 20, 89-109.

Valente, T.W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How Correlated Are Network Centrality Measures? *Connections,* 28, 16-26.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

Reproductive Health Policy and Interstate Influence

Lawrence B. Finer
Guttmacher Institute, New York, NY

Nan Marie Astone
Department of Population and Family Health Sciences
Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD

Thomas W. Valente
Department of Preventive Medicine
Keck School of Medicine, University of Southern California, Alhambra, CA

This paper compares two models of interstate influence: the proximity model (which posits that states are more influenced by nearby states than farther ones) and the opinion-leader model (which hypothesizes that some states are "regional leaders" which exert a disproportionate influence on all other states). These models are compared for two outcomes, reproductive health policies and general liberalism, using regression models enhanced by network analysis techniques. Data on the 50 states were collected. Dependent variables included policies and spending on reproductive health as well as a broader range of policies designed to measure general liberalism. Independent variables included historical and geographic conditions, socio-economic factors, political behavior, governmental institutions, and the behavior of elites. Results indicate that a state's policies on reproductive health (excluding abortion) appear to be a function of both the socio-economics of the state as well as the reproductive health policy of its regional leaders. General liberalism towards reproductive health, in contrast, is largely explained by per capita income and percentage of the state population that is fundamentalist, as well as the liberalism of a state's geographic neighbors. The importance of acknowledging network effects in state analyses, and of progressive states' leadership in advancing reproductive health, is underscored.

Authors

<u>Lawrence B. Finer</u> *is Director of domestic research at the Guttmacher Institute, overseeing projects on family planning services, contraceptive use, pregnancy and abortion, and adolescent reproductive health. He also serves as a senior lecturer at Columbia University's Mailman School of Public Health.*

<u>Nan Marie Astone</u> *is an Associate Professor and the Director of graduate education in the Department of Population, Family and Reproductive Health at the Johns Hopkins Bloomberg School of Public Health. She is also a member of the core faculty at the Hopkins Center for Adolescent Health.*

<u>Thomas W. Valente</u> *is a Professor and Director of the Master of Public Health Program in the Department of Preventive Medicine, Keck School of Medicine, University of Southern California. Dr. Valente uses network analysis, health communication, and mathematical models to implement and evaluate health promotion programs.*

Correspondence: *Contact Lawrence B. Finer, Director of Domestic Research, Guttmacher Institute, 125 Maiden Lane, New York, NY 10038. Phone: (212) 248-1111; fax: (212) 248-1951; email: lfiner@guttmacher.org*

INTRODUCTION

One of the main assumptions of regression analysis is that each observation of an outcome variable being studied is independent from the others (Weisberg 1985). In a study where the unit of analysis is the state, however, this assumption is certainly violated (Doreian 1980). State governments do not operate in isolation, and laws may develop partly in response to conditions in other states. The goal of this paper is to discuss possible forms of interstate influence and to test for the impact of those influences in one specific policy arena — public spending and lawmaking around reproductive health — using regression models enhanced by network analysis techniques.

1. Interstate Influence and Network Analyses

In what ways can states influence each other? Simple proximity is one possibility. The fifty states have been described as policy laboratories where federal programs are implemented and new program approaches are invented and tested, and states may view their neighbors as experimental laboratories (Elazar 1972; Gray 1990). Although state policy actions may be the result of independent invention or innovation, it is often the case that a new policy will diffuse across the states, many times without federal government action (Savage 1985). In other words, state legislatures may be influenced by actions nearby states have taken. Work on state policy innovation (Berry and Berry 1990) found empirical support for neighboring states' influence in a study of state lottery adoptions. Proximity, then, can be considered a measure of interstate influence, and a state may be most susceptible to the actions of contiguous states.

It is also possible that state policymakers may be influenced by states that are "opinion leaders" — in other words, states that are at the forefront of policy adoption or whose actions garner significant national attention. Walker's (1969) analysis of the diffusion of innovation among the states described the importance of regional leaders and the role of associations of state legislatures, and Strang (1996) suggested that regional influence varies by state size, with

smaller states being more susceptible to adopting the behaviors of bigger states.

Spatial and opinion-leader influences may be slightly contrasting, since one (the opinion-leader effect) involves specific roles, while the other is a broader, more global characteristic. Measuring these influences is a task suited to the techniques of social network analysis (SNA) SNA is a method of analyzing communication in a social system by determining who is in contact with whom and who influences whom (Wasserman and Faust 1994; Scott 2000; Valente and Foreman 1998). While often examined in the context of the diffusion of individual-level behaviors, network analysis of aggregate geographical units has also been the focus of some research. As such, it is often referred to as spatial-effects analysis, since the work emphasizes the spatial interrelationships between observations that are not captured by the characteristics of each individual observation (Ord 1975; Loftin and Ward 1983). Diffusion of behavior (or policymaking) can thus be approached with a temporal focus emphasizing changes in the policy "over time within a given spatial unit," or with a spatial focus, which "is more concerned with the extent to which adoption of the behavior (or policy) has traversed spatial boundaries by a given point in time" (Tolnay 1995). This paper compares the proximity and opinion-leader models of interstate influence for two outcomes, reproductive health policies and general liberalism.

2. Reproductive Health in the States

Rates of unintended pregnancy and abortion as well as contraceptive "method mix" — the percentage of a population using each of various methods — vary significantly across Western nations. While intended-pregnancy rates are quite consistent, unintended pregnancies are much more common in the United States than in Canada, Britain and most other developed nations, as evinced by the substantially higher abortion rate in the U.S. than in most of these countries (Coleman 1983; Brown and Eisenberg 1995). Similarly, other Western nations exhibit a markedly different mix of methods than that of

the United States; while sterilization and the pill predominate in the U.S., the IUD and the condom are used by a significant proportion of several non-U.S. populations (United Nations 2008).

These differences cannot be wholly explained by differences in demographic composition, for the assumption that all methods are equally available in each nation is simply not accurate. For example, lack of demand cannot explain the almost complete absence of the pill, the IUD and sterilization from the Japanese contraceptive repertoire. It has been suggested that the long delay in approval of the pill in Japan (in 1999) was partly due to the medical profession's fear that more effective methods would reduce the need for lucrative abortion services (Jitsukawa and Djerassi 1994, Associated Press 1999).

The different patterns of national family planning utilization may therefore reflect different supply-side factors operating in each country. In most countries, these factors tend to operate at the national level. In the United States, however, the largest American states are comparable in population size to several of the smaller European countries, and the federalist political structure make these states more like miniature semi-independent nations (Haub and Yanagishita 1994; Gray and Jacob 1996). Political-sociological analysis of social policy in America has generally disregarded this point, proceeding instead from a traditional European model that emphasizes the importance of nation-state policies and the role of the central government in social provision. However, as Skocpol (1992) points out, "The United States has never come close to having a 'modern welfare state' in the British, the Swedish, or any other positive Western sense of the phrase." The nationalized medical care model prevalent in France, Britain and other major European nations does not apply to the United States. Rather, the American pattern is distinguished by the lack of a national health care program and a wide range of health care delivery structures (Gold and Richards 1996).

The same bias toward a centralized perspective has been demonstrated in analyses of

reproductive health care. The legality of contraception in this country has been nationally uniform since the 1965 *Griswold v. Connecticut* decision, in which the U.S. Supreme Court found that married persons have a constitutionally protected right to use contraception. Subsequent Supreme Court decisions in 1972 (*Eisenstadt v. Baird*) and 1977 (*Carey v. Population Services International*) expanded this right to the unmarried and to minors (Hall et al. 1992). Because of this uniformity, the analysis of supply-side influences in the United States has focused on national factors, looking to the contraceptive development and regulatory environments as primary determinants of use.

However, the influence of decentralized social service provision has been felt particularly strongly in the area of reproductive health. The U.S. does not have (and has never had) an official national family planning policy (Brown and Eisenberg 1995), and this nation "differs from the usual pattern in that contraceptive care is not offered to everyone at little or no expense and, like most [U.S.] health care, is delivered primarily through medical specialists" (Jones et al. 1988: 65). Despite the existence of a federally funded family planning program (the Title X program, established by Congress in 1970), federal-state collaborations, particularly Medicaid, have come to dominate public funding of reproductive health, and state-level actions can have enormous influence over a range of policies that determine who receives reproductive health care, including contraception, abortion, and prenatal care, particularly for low-income populations. Indeed, state policies on reproductive health spending and restrictions on access to services vary widely by state (Alan Guttmacher Institute 2004).

Some analysts have examined reproductive health policymaking in the states, but primarily at a later stage of the policy process: implementation. In a series of articles, McFarlane (1983; 1985; 1989; Meier and McFarlane 1996) tested one tenet of Mazmanian and Sabatier's (1989) model of policy implementation, namely that more "coherent"

statutes — ones with clearer goals and implementation paths — are more effective. They compared the various federal funding streams and found that Title X dollars had a greater impact than other sources of federal funding. Although this research reinforces the finding that federal dollars are often more effective than state funds are at achieving public health goals, these studies do not provide insight into the relative importance of historical, economic and political factors in state-level policy construction.

It is important to note that in the area of reproductive health, interstate influence may be particularly salient. This is because individuals seeking reproductive health services are more likely to see contiguous states, rather than states in a region, as potential sources of services; thus policymakers are forced to be conscious of the regulations in effect in neighboring states.

METHODS

1. Operationalizing Network Effects

A general model for network-effects analysis is similar to the general linear model, but adds a term to account for network influence (Erbing and Young 1979; Valente 1995; Burt 1987). Such a model can be specified as:

$$y = \mathbf{X}\beta + \mathrm{NE}\gamma + \varepsilon \qquad (1)$$

where y is a vector of n observations on a dependent variable of interest, $\mathbf{X}$ is an n-row by k-column matrix of n observations on k fixed predictor variables, β represents the vector of regression coefficients for the k predictor variables, NE is an n-observation predictor variable representing the network effect (which has not yet been defined) of surrounding geographic units (i.e., states), γ is the (scalar) coefficient representing the impact of the network effect variable, and ε represents the vector of n random errors in the model. If γ is significantly different from zero, then some sort of network effect exists (Land and Deane 1992; Roncek and Montgomery 1984).

How is the network-effect variable constructed? In essence, a model that includes network or spatial effects implies that a geographic unit's score on a particular outcome variable (e.g., supportiveness of reproductive health policy) is affected by the values of other, "nearby" geographic units on that outcome variable. In order to measure network influence, therefore, one must construct a variable that summarizes those "nearby" values, either by summing them, averaging them, or weighting them by some "distance" or "proximity" factor:

$$\mathrm{NE}_i = \frac{\sum_{j \neq i}^{50} \mathrm{Y}_j \left(w_j^i \right)}{\sum_{j \neq i}^{50} w_j^i} \qquad (2)$$

Here, NE_i is the network-effect score for state i, Y_j is the outcome-variable value for the jth state, w_j^i is the weight (as yet undefined) for state j in relation to state i, and the summation is taken over all states j other than i.

2. Modeling the Effect of Distance

The key question, then, is how to operationalize the network-effects weights. In the simplest conception of network effects, every other state would have equal impact (i.e., a weight of 1), and one would simply average the scores of all other states. However, such a measure would not account for two essential components of network effects. First, an important assumption is that distances between states matter; closer states should have more influence than more distant states. Second, expanding on Strang's observation that smaller states are more susceptible to network effects, there may be some sort of "threshold effect" — a qualitative difference between the influence of immediately surrounding states (i.e., contiguous states) and states that are even one step removed. Consider the state of Idaho, for example. One may theorize that Utah or Wyoming (each contiguous to Idaho) has a stronger effect on that state than, say, Illinois, and one may also suspect that both Illinois and Virginia are equally non-influential, despite the substantial difference in their

distances from Idaho. In this way, smaller states and those whose neighbors are generally smaller would have more neighbors falling within the threshold and thus would be more susceptible to influence. (It is interesting to note that this contrasts with the previously discussed suggestion that small states are more susceptible to network influences than bigger states.)

Roncek and Montgomery (1984) propose weighting each surrounding state's score by the inverse of the distance between it and the state in question:

$$w^i_j = \frac{1}{d_{ij}} \qquad (3)$$

Because the weighting factor is the inverse of the distance between states, the weights fall off very quickly as distance increases. Thus, the effect of nearby states will be emphasized, and that of faraway ones will be de-emphasized. Although this is desirable, the falloff is very steep for any but very small distances (see Figure 1).

Another approach is to use the *reverse* rather than the *inverse* as the metric for weighting (Valente and Foreman 1998). The resulting metric is referred to as radiality. In this approach, each interstate distance is subtracted from the largest interstate distance to obtain a measure of proximity:

$$w^i_j = \frac{d_{\max} - d_{ij}}{d_{\max}} \qquad (4)$$

This approach has the advantage of maintaining the original metric of the distances; in other words, proximities are directly proportional to the original distance measures. The influence falloff is not very steep for this approach (Figure 1). In order to model the threshold effect described above, one might apply an explicit threshold criterion: One might "zero out" scores for noncontiguous states, or for states that are greater than a certain distance (say, 500 miles)

away. The formula for the network-effects weight would then be:

$$w^i_j = \frac{d_{\max} - d_{ij}}{d_{\max}}, \quad d_{ij} \le 500$$
$$w^i_j = 0, \quad d_{ij} > 500 \qquad (5)$$

Ideally, however, one would be able to model the effect of distance more elegantly, through a formula that allows influence to fall off slowly at first, then more sharply after a certain distance, after which influence would be minimal. A descending logistic function provides such values and can be generated by the equation:

$$w^i_j = \frac{\exp\left(a[d_{ij}] + b\right)}{1 + \exp\left(a[d_{ij}] + b\right)} \qquad (6)$$

where exp represents the base of the natural logarithm and *a* and *b* are parameters that determine the shape of the curve. Figure 1 demonstrates the shape of the logistic curve with parameters *a* and *b* intentionally set at -0.015 and 8, respectively, as well as the reverse and inverse curves, for distances of 0 to 2000 units.

The logistic function satisfies the threshold criterion and allows for a reasonable rate of falloff for values above the threshold (i.e., distances of 0 to 500 units). Thus, in the present analysis, the logistic function in equation 6 was used to calculate network weights, and then equation 2 was used to calculate the network effects for specific outcome variables. In practice, the variables were calculated by (1) creating a 50×50 matrix of city distances (i.e., between state capitals), (2) applying the logistic function to each cell to produce a matrix of network weights, (3) multiplying these weights by the 50×1 vector of values for each outcome variable, and (4) dividing each value by its row sum.

Figure 1. Modeling Interstate Weights
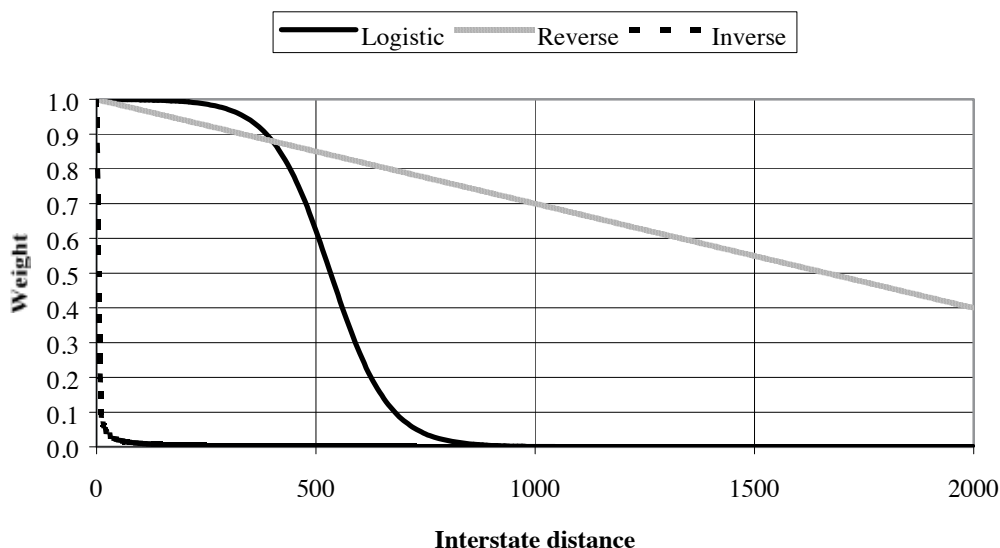
3. Modeling Regional Leadership

The weighting strategy discussed captures the distance component of interstate influence. But another goal of the analysis was to account for a second form of interstate influence, Walker's (1969) "regional leadership" hypothesis described above, where states are recognized as influential because of size, wealth or a reputation for being at the forefront of policymaking. A regional leader was identified in each of the nine Census subregions, and a network effect variable was created from the score of that state's regional leader on each outcome.[*]

Following are the nine Census subregions and the states in each (the regional leader is bolded):

New England: **MA,** CT, ME, NH, RI, VT

Mid-Atlantic: **NY,** NJ, PA

East North Central: **MI,** IL, IN, OH, WI

West North Central: **MN,** IA, KS, MO, NE, ND, SD

South Atlantic: **MD,** DE, FL, GA, NC, SC, VA, WV

East South Central: **KY,** AL, MS, TN

West South Central: **LA,** AR, OK, TX

Mountain: **CO,** AZ, ID, MT, NV, NM, UT, WY

Pacific: **CA,** AK, HI, OR, WA

4. Dependent Variables

Two indices were created and used as dependent variables in this analysis. These indices resulted from a factor analysis of a variety of policy and spending variables at the state level, including both reproductive health-specific and general policies. Variables represented the policies that were in effect or the spending that took place during or close to 1995. The policies included in the factor analysis loaded onto two factors; the first contained virtually all the reproductive health policies, while the second contained general policies.

The first measure can be seen as an index summarizing reproductive health spending and policymaking at the state level. This index excludes measures of abortion policy, because

---

[*] Alaska and Hawaii were not included in Walker's analysis, but were added to the Pacific subregion.

the factor analysis findings indicated that abortion is more appropriately analyzed separately from other reproductive health policy outcomes. The reproductive health index is useful for testing the hypothesis that a variety of historical, environmental, socioeconomic and governmental factors (see the next section), both related and unrelated to reproductive health, are associated with reproductive health policy outcomes.

The second index represents a more general measure of policy liberalism. Using this second index allows one to compare a state's reproductive health orientation to a general liberal/conservative characterization of each state. Most studies of state adoption of reproductive health policy would expect that political orientation is highly correlated with, and indeed causes, state reproductive health policy, and that liberal states would therefore support reproductive health policies more strongly than conservative ones.

The following components are included in each factor:

4.1 Reproductive health factor

- Contraceptive spending per woman in need of publicly funded reproductive health services
- Minors can consent to prenatal care (y/n)
- Minors can consent to contraceptive services (y/n)
- State has an extended postpartum stay law (y/n)
- Medicaid income cutoff for pregnant women
- Simplified Medicaid application for pregnant women (y/n)

4.2 General liberalism factor

- Pregnant women with no children are eligible for Aid to Families with Dependent Children (AFDC) (y/n)
- Medicaid expenditure per beneficiary
- Index of commitment to environmental protection

- AFDC spending per recipient
- State permits death penalty (y/n)
- State prohibits sodomy (y/n)

The reproductive health score coefficients indicated that each of the six component variables contributed about the same amount to the overall reproductive health factor. This factor score itself was therefore used as a summary measure for the analyses. However, the liberalism factor weighted more heavily toward those variables related to spending, so rather than use the factor scores, a liberalism index was created by standardizing and then summing the six variables.

5. Independent Variables

A modified version of Hofferbert's 1974 systems model for public policy analysis was used to guide the choice of independent variables. In Hofferbert's "funnel" model, five sets of factors affect a progressively narrower set of issues. He sees (1) historical and geographic conditions as the widest environmental influence, affecting (2) socioeconomic conditions, which in turn affect (3) mass political behavior (including voter turnout and party structure), (4) governmental institutions (including legislative and administrative bodies), and (5) the behavior of governmental elites. All of these factors contribute to the production of policy outputs, and each factor works directly and through subsequent factors to exert influence.

It can be argued that the broadest category in the framework should reflect historical-political as well as general historical conditions. In the models, this level of the framework was represented by the structure of state family planning delivery systems (i.e., whether state Title X dollars were overseen by a government entity such as a health department or an independent entity such as a Planned Parenthood affiliate). Also included was Elazar's 1966 measure of "political culture," categorizing states as moralistic (in which the political process is seen as an active tool for improving society), traditionalistic (where an entrenched political system is led by an insular, elite group

of politicians), or individualistic (where a more libertarian outlook predominates).

Socioeconomic conditions were included through conventional measures of state income, wealth, urbanization, educational achievement and racial/ethnic composition. Measures of mass political behavior included public opinion liberalism and political party membership, two classic indicators of political representation. To operationalize governmental institutions, measures of legislative professionalism — specifically, Squire's (1992) index comparing state legislatures to the U.S. Congress in terms of pay, staff support, and time spent in session — were included, as well as female representation in state legislatures.

Finally, influences operating at the level of political elites were represented by the strength of reproductive health and religious advocacy groups, as well as state governors' support for reproductive health. Also included was a measure of issue salience: the number of bills related to reproductive health introduced in the state legislature.

6. Univariate, Bivariate and Multivariate Analyses

Initially, the two indices alone were examined in order to assess them substantively, and then cross-plotted to look for correlation. The indices were also mapped to look for geographic patterns. Because of the relatively small number of observations and the large number of independent variables under consideration, a series of smaller regressions were performed first in which each outcome variable was regressed on each individual predictor variable, and then on each subset of related predictors (e.g., socioeconomic predictors) as a group. By doing so, a first-stage assessment was made of which variables might be significant in a larger multiple regression model. Numerous regressions were then performed using these potentially significant variables in many combinations, primarily examining the resulting coefficient significance levels and model R^2 values (or, in the case of logistic regression,

pseudo-R^2 values). Several of the more promising models were adjusted to examine the role each variable played in a changing context. The goal of doing so was to find models that explained a high proportion of the variance in the outcome variable with as few predictors as possible.

Once this second reduction was completed, several interaction terms were added to the models, and variables that had made the first but not the second "cut" were reintroduced. Significant interaction coefficients and coefficients with signs opposite those of the main effects were noted, as were interactions that resulted in a substantial increase in the model's R^2 value. Finally, after an optimal model was established for each dependent variable, each model was adjusted again to include the network-effects variables.

RESULTS

1. Reproductive Health and Liberalism Indices

Table 1 lists the reproductive health index and the standardized index of liberalism for each state, in descending order. The most noticeable trend among the reproductive health index is the appearance of several Southern states (Maryland, Virginia, Kentucky, North Carolina) at or near the top of the list. A number of states generally considered progressive, such as Massachusetts, New York and Hawaii, appear high on the list as well. At the bottom of the list are several Plains states, such as Nebraska, Utah and the Dakotas, as well as some Northeastern states like Pennsylvania and Ohio.

The list for the liberalism index is somewhat different. Massachusetts, Minnesota and New York again appear high on the list, but some states that scored low on reproductive health, such as North Dakota and New Hampshire, score high on the liberalism index. Southern states predominate in the bottom portion of the list.

Table 1. Reproductive Health Factor Scores and Liberalism Index

Reproductive health		Liberalism	
State	Factor score	State	Index value
Maryland	1.47	Massachusetts	6.86
Minnesota	1.29	Wisconsin	5.78
Virginia	1.27	New York	5.47
New Mexico	1.22	Minnesota	5.23
Washington	1.22	Alaska	5.20
Kentucky	1.20	Connecticut	4.79
North Carolina	1.19	North Dakota	4.27
Massachusetts	1.11	Maine	4.24
New York	0.98	Vermont	4.17
Hawaii	0.98	Hawaii	3.15
South Carolina	0.97	Nebraska	3.02
Tennessee	0.87	Rhode Island	2.30
Florida	0.87	Delaware	2.30
New Jersey	0.65	New Hampshire	2.29
California	0.56	New Jersey	2.22
Arkansas	0.52	Iowa	2.11
Delaware	0.52	Oregon	2.10
West Virginia	0.46	California	1.79
Mississippi	0.45	Wyoming	1.68
Michigan	0.21	Michigan	1.65
Georgia	0.18	Ohio	1.42
Colorado	0.07	Pennsylvania	1.30
Illinois	-0.04	Washington	1.02
Alaska	-0.15	Colorado	0.44
Alabama	-0.17	Indiana	0.37
Oklahoma	-0.20	Illinois	0.12
Texas	-0.24	New Mexico	-0.01
Oregon	-0.27	Utah	-0.06
Kansas	-0.31	North Carolina	-0.40
South Dakota	-0.34	Idaho	-0.50
Vermont	-0.37	Nevada	-0.51
Maine	-0.39	Maryland	-0.57
Missouri	-0.43	Kansas	-0.96
Montana	-0.45	Arizona	-1.18
Arizona	-0.51	Montana	-1.41
Indiana	-0.57	Florida	-1.63
Wisconsin	-0.62	South Dakota	-1.65
Idaho	-0.69	West Virginia	-2.34
Iowa	-0.75	Tennessee	-3.74
Rhode Island	-0.77	Louisiana	-3.89
Connecticut	-0.81	Virginia	-3.99
Louisiana	-0.91	Oklahoma	-4.03
New Hampshire	-0.96	Missouri	-4.87
Ohio	-0.99	South Carolina	-4.90
Utah	-1.02	Georgia	-5.30
Pennsylvania	-1.02	Kentucky	-5.35
Wyoming	-1.08	Texas	-6.17
North Dakota	-1.17	Alabama	-6.59
Nebraska	-1.53	Arkansas	-6.60
Nevada	-1.53	Mississippi	-8.68

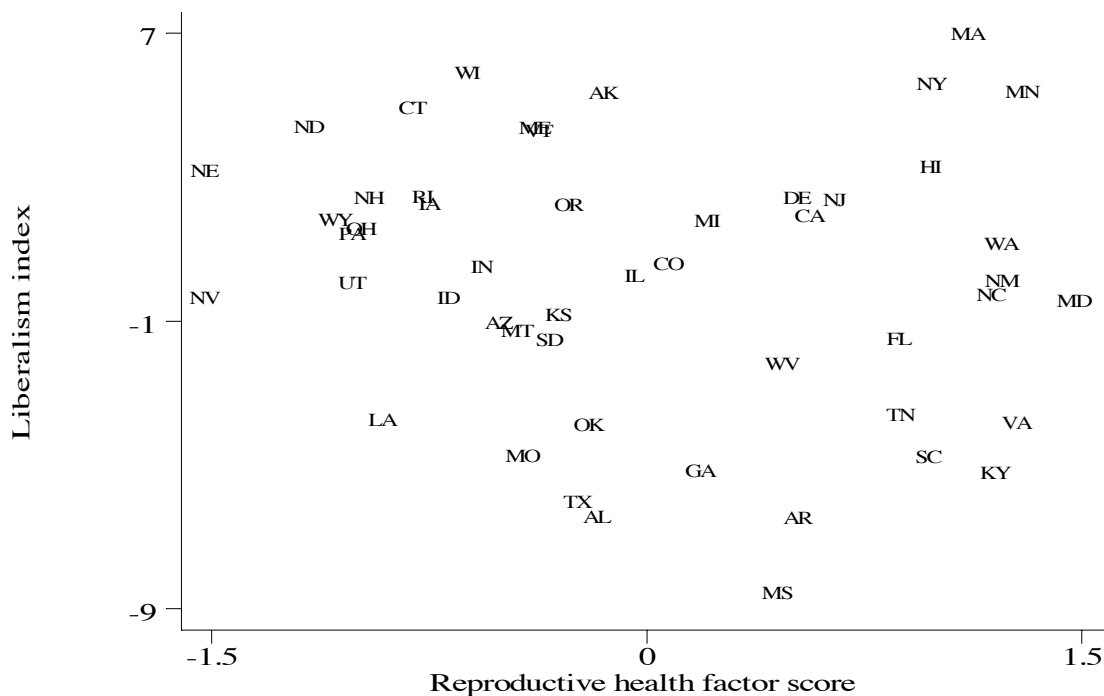**Figure 2. Scatterplot of Reproductive Health and Liberalism Scores**

Figure 2 facilitates this comparison by graphically cross-plotting the reproductive health and liberalism scores for each state.

If states' relative scores were similar for both factors, one would expect to see the states falling on a diagonal line from the lower left to the upper right corner of the graph, but this is clearly not the case. The most notable tendency occurs in the bottom right corner of the scatterplot. Southern states such as Kentucky, South Carolina, Virginia, Arkansas, Mississippi and Tennessee score in the bottom half of the liberalism scale but in the top half of the reproductive health scale. Other Deep South states fall in this area as well.

New York, Massachusetts and Minnesota are grouped together in the top right corner, suggesting that they are the most consistently liberal states on both dimensions. In the upper left corner, one sees a number of Plains states like Nebraska and North Dakota, along with Northeastern states like Connecticut, New Hampshire and Pennsylvania, as well as Wisconsin.   This position indicates high liberalism but low reproductive health support. In the lower left corner is Louisiana, receiving low scores on both dimensions.

Figures 3 and 4 show the relative factor scores for reproductive health and the liberalism index. In each case, darker shading indicates more support for the respective policy category. In comparison to the more patchwork appearance of colors on the reproductive health map, the general liberalism map shows more continuous gradations, suggesting that geographic proximity may be found to be more relevant for this outcome variable.

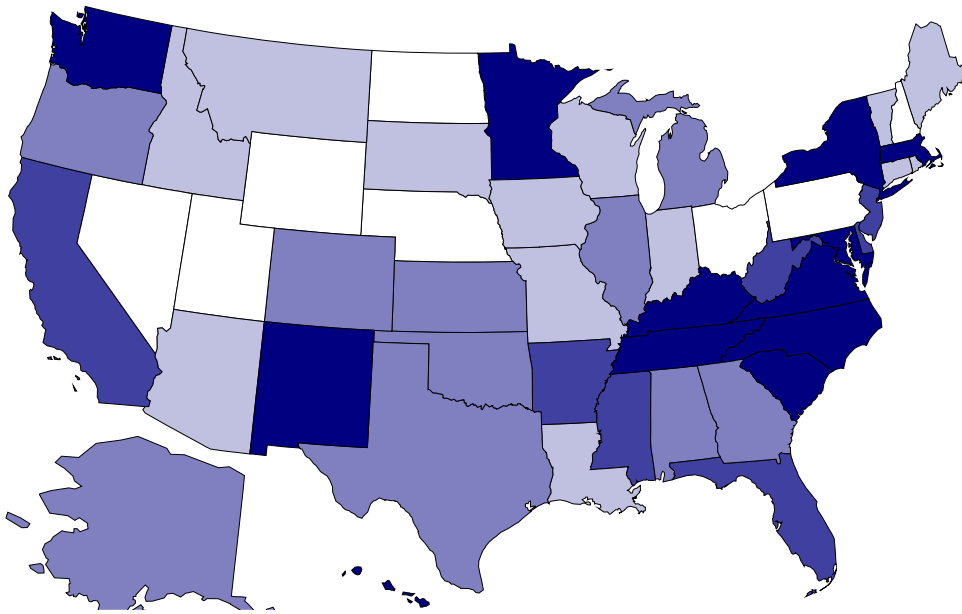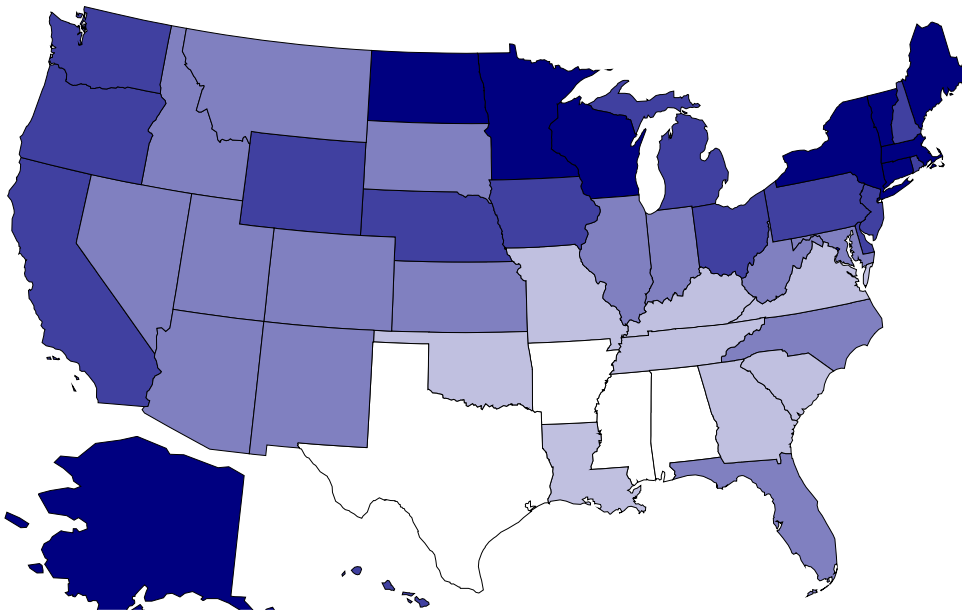**Figure 3. Mapping of Reproductive Health Factor Scores**
(darker = more supportive)

Figure 4. Mapping of Liberalism Index
(darker = more liberal)

2. Regression Models

The simple regressions on the broad list of possible independent variables produced the following list of predictors that had potential explanatory power:

- Year abortion legalized
  (if before *Roe v. Wade*)
- Southern state
- Median income
- Percent in poverty
- Traditionalistic political culture
- Percent with a high school diploma
- Percent with a college degree
- Percent minority
- Percentage of the population that
  is fundamentalist
- Legislative professionalism
- Percent of state legislators that
  were Democratic
- Percent voting for Ross Perot in 1996
- Number of reproductive health bills
  introduced or acted on
- Number of NARAL members
- Conservative group affiliates per capita
- Governor's stance on abortion

Some of the null findings – variables that were not significant – are of interest. Most notable is public opinion liberalism, which was not significant even when the reproductive health factor was regressed on this predictor alone (not shown). Neither Fundamentalist nor Catholic representation was associated with reproductive health policy support, although my measure of conservative affiliates and the measure of NARAL membership were both at least marginally significant. Interparty competition, prominant in the literature to date, also did not demonstrate an association with reproductive health policy.

The variables listed above were used in a second set of regressions, which attempted to "fine-tune" the model. In doing so, it was found that several variables maintained their significant status when entered into the model at the same time. The inclusion of seven predictor variables produced an R^2 value as high as .59, but the incremental gain brought by each additional variable was minimal. The addition of interaction terms, however, resulted in R^2 values as high as .71. Model 1 in Table 2 produced the best fit.

Once that model was settled on as the "optimal" model, the network effects that were previously calculated were added. Recall that the "distance" effects were based on the distance between state capitals, and the "regional leader" effects were constructed by choosing a handful of "regional leaders" and then identifying which states were geographically close to those regional leaders. Models 2 and 3 in Table 2 shows revised regression coefficients including the two network variables.

The models that follow (Table 2) offer substantial evidence that Southern states differ from the rest of the nation both in their support of reproductive health and in the factors that affect that support. The large positive coefficient indicates that Southern states are more supportive of reproductive health in general, and the included interaction terms shed light on how that support plays out. Both poverty and education are positively associated with reproductive health support. However, the coefficient for the South × poverty interaction term is negative and larger than the poverty coefficient, indicating that in the South, the effect of poverty is reversed: A larger poor population is associated with a *lower* (i.e., more conservative) reproductive health policy score. Despite the higher reproductive health support scores in the South in general, poorer Southern states provide less support for reproductive health.

In Model 3, the regional-leader effect variable is significant, indicating that this form of interstate influence is relevant for reproductive health. Inclusion of this variable did not absorb the effect of other predictors; virtually all of the predictor variables that were significant in Model 1 remained significant in Models 2 and 3, although most coefficients became slightly smaller, and some dropped down one level of significance. The R^2 of the model rose to .75.

Table 2. Reproductive Health: Final Regression Models

Variable	Model 1	Model 2	Model 3
Southern state	8.29***	8.11***	6.58**
Percent in poverty (log transformed)	1.16*	0.975*	0.994*
South × log(poverty)	-2.02**	-1.93**	-1.37†
Percent minority (square-root transformed)	-0.0300	-0.0178	0.0566
South × sqrt(minority)	-0.376**	-0.377**	-0.375**
Percent with high school diploma	0.0830**	0.0741*	0.0794**
Percent voting for Perot in 1996	-0.142**	-0.141**	-0.112*
Number of RH-related bills introduced	0.0286**	0.0278**	0.0236**
Governor's support of abortion	0.292**	0.305**	0.287**
Distance network effect	-	0.237	-0.360
Regional-leader network effect	-	-	0.259*
Constant	-9.32**	-8.21*	-9.39**
N	50	50	50
R^2	0.71	0.72	0.75

†$p<.10$, *$p<.05$, **$p<.01$, ***$p<.001$

Table 3. Liberalism Index: Final Regression Models

Variable	Model 1	Model 2	Model 3
Southern state	-3.49***	-2.78*	-
Median income per capita	0.246***	0.209**	0.225**
Percent Fundamentalist	-0.081**	-0.064*	-0.0797*
Does state have at least one Democratic chamber?	1.71*	1.56*	0.636
Percent of population voting for Perot in 1996	0.477**	0.376*	0.425*
Distance network effect	-	0.286	0.568**
Regional-leader network effect	-	0.0380	-
Constant	-11.8***	-9.92**	-10.96**
N	50	50	50
R^2	.78	.79	.74

*$p<.05$, **$p<.01$, ***$p<.001$

In regressions on the liberalism index (Table 3), a model including just five predictor variables (other than the network variables) accounted for more than three-fourths of the variance in the outcome variable.

Unlike reproductive health, Southern states are less likely to have a high liberalism score, which might be expected. Indeed, the South is nearly synonymous with traditionalistic political culture; only two Southern states (Delaware and Maryland) are not described as traditionalistic by Lowi, and only two traditionalistic states (Arizona and New Mexico) are not technically in the South.

In Model 2, which includes the predictor "South," both of the network weight variables are insignificant. However, Model 3 shows that when South is removed, the distance weight is strongly significant, as Figure 4 suggests. In addition, the correlation between the liberalism outcome variable and the liberalism-specific distance weight is .78. Clearly, states are likely to have a general liberalism score that is similar to their neighbors' scores.

DISCUSSION

In this analysis, supportive reproductive health policy appears to be a function of a wide range of factors, including several socioeconomic variables. In addition to these, however, a state's reproductive health policy is also a function of the reproductive health policy of its regional leaders. General liberalism, in contrast, is largely explained by per capita income and percent fundamentalist. Importantly, general liberalism is a function of the liberalism of a state's neighbors.

The Southern support of reproductive health as defined in this analysis is a striking finding. Even with state poverty included in the model, the South was still strongly associated with reproductive health support. There may be several explanations for this support in the South. It may reflect efforts by Southern legislatures to save money, or a remnant of traditionalistic culture. The significant and positive association of poverty with reproductive

health support is in accordance with earlier findings that states with a poorer populace are more financially supportive of their citizens.

However, this finding is tempered by the negative association in the South between the size of the minority population and reproductive health support. Even after controlling for poverty, Southern states with higher minority populations — Texas, Mississippi, Louisiana — provide less public support for reproductive health than other Southern states.

Liberalism, as defined in this project, is a product of a more limited set of factors. Most of the key predictors of liberalism conformed to expectations, as Southern states and more Fundamentalist states were less liberal, and income and Democratic control were positively associated with liberal orientation.

In contrast to the above variables, the importance of the Ross Perot vote is one of the most unexpected findings to come out of this project. Perot's independent third-party candidacy for president in 1996 espoused both liberal and populist positions, including electronic "town halls," trade protectionism and, interestingly, support for abortion rights. It was therefore unexpected that having a high percentage of Perot voters would be negatively associated with reproductive health support (although it did have a positive association with general liberalism). This variable correlates positively both with spending variables such as AFDC and with social policy variables such as sodomy legality and death penalty prohibition (although it does not correlate highly with environmental protection). This correlation is likely driven primarily by regional effects: Support for Perot was strongest in northern Plains states such as Montana, Idaho, Wyoming and North Dakota, as well as the far Northeastern states of Maine, Vermont and New Hampshire, and these two areas are not bastions of reproductive health support, despite scores on the liberalism index sufficient to allow a positive association there. The finding suggests a disconnect between the perhaps more libertarian character of these states and aspects of Perot's

platform supporting government involvement in many issues.

In the reproductive health policy models above, interstate influence was shown to play a significant role in predicting these policies. Regional leaders appeared to have an influence on their surrounding states: A state's reproductive health support was significantly associated with the score of the opinion leader in that state's region. For liberalism, on the other hand, there was evidence of a proximity influence, both visually (see Figures 3 and 4) and statistically (see Table 2). The checkerboard pattern of the reproductive health map, compared to the more gradual changes in the liberalism map, suggests that reproductive health policy, contrary to expectations, is not as susceptible to neighborly influence; prominent states' actions may play a larger role.

Some methodological and substantive factors could call into question the findings in this study. Conclusions drawn from data are accurate only to the extent that (1) the data themselves are accurate and (2) the actual measurement accurately reflects the underlying construct one is attempting to measure. Some of the data, such as those on Medicaid funding limits for abortion and those on public opinion and legislative professionalism, are several years away from the focal year of the analysis (1995). Although many of these variables may be considered fairly stable, others, such as Medicaid limits, may change more frequently. The regional-leader weights may be another case in which these assumptions are questionable. The measure used is based on work from three decades ago, and state leaders and followers may have shifted since then. Even so, any attempt to measure this influence is an improvement over ignoring the concept entirely, and future work may benefit from additional efforts to refine these measures of influence.

The findings described here underscore both the peculiarity of the American model and the distinctiveness of reproductive health policy. Reproductive health occupies a unique niche between ordinary health care — which is actually "sick care" — and straightforward

social redistribution. Conservatives may not see it as a necessary part of health maintenance, so it is not automatically considered worthy of funding. On the other hand, it may reduce unintended births and abortions, so it can be seen as worthwhile from an economic perspective. Here, conservatives may be joined by libertarians, who might consider family planning as a tool to reduce economic dependency.

The finding that reproductive health policy is a product of many factors, historical and economic as well as political, is both encouraging and discouraging for those wishing to advocate more supportive policies in this arena. The broad range of predictors of the reproductive health factor implies that issue-focused advocates may have an easier time affecting one specific policy than an entire substantive area. If one assumes that policy does affect health outcomes, it may therefore be quite a challenge to improve reproductive health, broadly measured, in less supportive states.

At the same time, the Perot example and the importance of "Southern exposure" suggest that support for reproductive health may exist where one least expects it. And if it is true that regional leaders can influence their peers, then it is essential for prominent states — many of whom are progressive in this area — to continue to lead the way in their support of reproductive health and family planning programs through strong policy measures.

REFERENCES

Alan Guttmacher Institute. 2004. State policies in brief [website]. <http://www.guttmacher.org/pubs/spib.html>.

Associated Press. 1999, June 2. Japan approves birth control pill. New York Times.

Berry, Frances Stokes and William D. Berry. 1990. State lottery adoptions as policy innovations: An event history analysis. American Political Science Review 84(2): 395-415.

Brown, Sarah S. and Leon Eisenberg. 1995. The best intentions: Unintended pregnancy and the well-being of children and families. Washington, D.C.: National Academy Press.

Burt, Ronald S. 1987. Social contagion and innovation: Cohesion versus structural equivalence. American Journal of Sociology 92: 1287-1335.

Coleman, Samuel. 1983. Family planning in Japanese society: Traditional birth control in a modern urban culture. Princeton, N.J.: Princeton.

Doreian, Patrick. 1980. Linear models with spatially distributed data: Spatial disturbances or spatial effects? Sociological Methods and Research 9: 29-60.

Elazar, Daniel J. 1966. American federalism: A view from the states. New York: Thomas Crowell.

Elazar, Daniel J. 1972. American federalism: A view from the states (2nd ed.). New York: Thomas Crowell.

Erbing, Lutz and Alice Young. 1979. Individuals and social structure: Contextual effects as endogenous feedback. Sociological Methods and Research 7: 396-430.

Gold, Rachel Benson and Cory L. Richards. 1996. Improving the fit: Reproductive health services in managed care settings. New York: The Alan Guttmacher Institute.

Gray, Virginia. 1990. The socioeconomic and political context of states. (5th ed.). Glenview, Ill.: Scott, Foresman/Little, Brown.

Gray, Virginia and Herbert Jacob. 1996. Politics in the American states: A comparative analysis. (6th ed.). Washington, D.C.: Congressional Quarterly.

Hall, Kermit L., James W. Ely Jr., Joel B. Grossman, and William M. Wiecek. 1992. The Oxford guide to the Supreme Court. New York: Oxford.

Haub, Carl and Machiko Yanagishita. 1994. World population data sheet. Washington, D.C.: Population Reference Bureau.

Hofferbert, Richard I. 1974. The study of public policy. Indianapolis: Bobbs-Merrill.

Jitsukawa, Mariko and Carl Djerassi. 1994. Birth control in Japan: Realities and prognosis. Science 265: 1048-1051.

Jones, Elise F., Jacqueline D. Forrest, Stanley K. Henshaw, Jane Silverman, and Aida Torres. 1988. Unintended pregnancy, contraceptive practice and family planning services in developed countries. Family Planning Perspectives 20(2): 53-67.

Land, Kenneth C. and Glenn Deane. 1992. On the large-sample estimation of regression models with spatial- or network-effects terms: a two-stage least squares approach. Chapter 7 in Marsden, P. V. (ed.), Sociological Methodology (Vol. 22). San Francisco: Jossey-Bass.

Loftin, Colin, and Sally K. Ward. 1983. A spatial autocorrelation model of the effects of population density on fertility. American Sociological Review 48(1): 121-128.

Mazmanian, Daniel A. and Paul A. Sabatier. 1989. Implementation and public policy (2nd ed.). Lanham, Md.: University Press of America.

McFarlane, Deborah R. 1983. The implementation of federal family planning policy: An analysis of factors affecting state expenditures, fiscal years 1976-1981. Unpublished doctoral dissertation, University of Texas, Houston.

McFarlane, Deborah R. 1985. Are block grants more responsive to state health needs? The case of the federal family planning program (FY 76-81). Journal of Health and Human Resources Management 8(2): 147-167.

McFarlane, Deborah R. 1989. Testing the statutory coherence hypothesis: The implementation of Federal family planning policy in the states. Administration and Society 20(4): 395-422.

Meier, Kenneth J. and Deborah R. McFarlane. 1996. Statutory coherence and policy implementation: The case of family planning. Journal of Public Policy 15(3): 281-298.

Ord, Keith. 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70: 120-126.

Roncek, Dennis W. and Andrew Montgomery. 1984. Spatial autocorrelation: diagnoses and remedies for large samples. Paper presented at the annual meetings of the Midwest Sociological Society, Des Moines, Ia.

Savage, Robert L. 1985. When a policy's time has come: Cases of rapid policy diffusion, 1983-1984. Publius 15(3): 111-125.

Scott, John. 2000. Network analysis: A handbook. 2nd ed. Newbury Park, CA: Sage.

Skocpol, Theda. 1992. Protecting soldiers and mothers: The political origins of social policy in the United States. Cambridge, Mass.: Belknap/Harvard.

Squire, Peverill. 1992. Legislative professionalization and membership diversity in state legislatures. Legislative Studies Quarterly 17(1): 69-79.

Strang, David. 1996. Inducing a network influence structure from multiple diffusion processes. Paper presented at the annual meetings of the American Sociological Association, New York.

Tolnay, Stewart E. 1995. The spatial diffusion of fertility: a cross-sectional analysis of counties in the American South, 1940. American Sociological Review 60(2): 299-308.

United Nations. 2008. World contraceptive use 2007 [chart]. New York: United Nations.

Valente, Thomas W. 1995. Network analysis of the diffusion of innovations. Cresskill, N.J.: Hampton Press.

Valente, Thomas W. and Robert K. Foreman. 1998. Integration and radiality: measuring the extent of an individual's connectedness and reachability in a network. Social Networks 20: 89-105.

Walker, Jack L. 1969. The diffusion of innovations among the American states. American Political Science Review 63: 880-899.

Wasserman, Stanley and Katherine Faust. 1994. Social network analysis: Methods and applications. Cambridge, UK: Cambridge University Press.

Weisberg, Sanford. 1985. Applied linear regression. New York: Wiley.

The Importance of Work-Related Social Ties in Post-Soviet Russia: Co-worker personal support networks in St. Petersburg and Helsinki

Markku Lonkila
University of Helsinki, Finland

This study considers the extent to which work-related social ties function as a source of social support in Russian workers' personal networks. The topic is important since, in the case of unemployment or retirement, personal networks are central for the well-being and coping of Russians. In order to illustrate the nature of the Russian case, an explicit comparison between Russian and Finnish workers' personal networks is carried out. The results are in line with previous findings concerning the workplace as a source of social support in China.

The Importance of Work-Related Social Ties in Post-Soviet Russia: The Role of Co-workers in the
Personal Support Networks in St. Petersburg and Helsinki

CONNECTIONS

INTRODUCTION

1. Cross-cultural Research on Personal Networks

There is an obvious lack of cross-cultural comparative studies utilizing a clearly defined notion of personal network and network data on alters' interconnections. This lack is partly due to the complexities of organizing a cross-cultural network research project which, even in the case of comparing two countries only, requires both time, money, and international collaboration contacts. More importantly, when designing the data collection and interpreting the results, a significant amount of cultural competence regarding all countries included in the study is needed.

Given these difficulties, international comparisons often rely on large pre-existing comparative surveys such as the World Values Survey, International Social Survey Programme and European Social Survey, using the survey questions on respondents' social relations to describe their personal networks. Though valuable in many respects, these surveys do not include data on alters' interconnections and therefore do not enable an analysis of the personal network structure. Fischer and Shavit (1995, 132) conclude, for example, that the International Social Survey Programme, one of the most comprehensive comparative studies, permits researchers to compare respondents' dyadic ties, but does not allow for comparisons of networks. Moreover, a focused study of the data questionnaires in original languages may reveal inconsistencies both in translations and cultural categories used in different countries.

A relatively small body of comparative research on personal networks utilizing network methods and collecting data on network structures has found both similarities and differences between countries. Fischer and Shavit (1995, 143) for example, found that the Israelis' networks were significantly denser than the Americans' and four). The results are presented in the fifth section, with conclusions in the final section.

conclude that "societal structures and cultures can selectively affect particularities of personal life" (Fischer and Shavit 1995, 143). Grossetti (2007), on the contrary, noted marked convergences in network density, between the personal networks of the Toulousains in 2001 and the Californians in Fischer's original study of 1977-78. He interprets this convergence by, among other things, the relatively stable relational structure in industrialized countries.

Research on support networks outside industrialized countries lends credence to the idea of cross-cultural variation in personal networks. Adams et al. (2006, 366) maintain, for example, that it is "clearly inappropriate to assume that the meaning, structure and function of support networks in Mali would be similar to those found in Western settings" and Lai (2001, 73) notes that the expectations from Chinese adult children to provide both material and emotional support to their elderly parents are more intense than in many other cultures.

This article contributes to the area of cross-cultural network studies through a detailed comparison of post-socialist Russia with neighboring Finland, a Nordic welfare society. It compares workers' personal support networks in the two countries on the basis of case studies conducted in Helsinki in 2003 and St. Petersburg in 2000 utilizing the network questionnaire adapted from Claude Fischer's original research (1982). Unlike many studies of social support focusing on family and kin ties, the study pays particular attention to the role of co-workers in the personal support networks.

The next section discusses the converging results of the studies of personal networks in Russia and China, both countries with experiences of the socialist system. The remaining text of the article focuses on the comparison between Russia and Finland, depicting data collection sites in St. Petersburg and Helsinki (section three), and the data and methods of the study (section

CONNECTIONS

The Importance of Work-Related Social Ties in Post-Soviet Russia: The Role of Co-workers in the
Personal Support Networks in St. Petersburg and Helsinki

2. The Significance of Work-related Ties as Source of Social Support in (Post) Socialism

Network research conducted in Russia and China has produced converging results on the importance of the work-related ties in both societies, relating this convergence to the legacy of the socialist era. Lonkila (1998), for instance, found in his comparison of 40 St. Petersburg and 38 Helsinki teachers in 1993 that whereas only 28% of all personal network ties of Helsinki teachers were mediated by their workplace, the corresponding figure for their St Petersburg counterparts was 48%. In a replication study with 20 teachers and five psychologists conducted in St. Petersburg in 1996, the same trend emerged even more clearly: work-mediated relations accounted for 53% of the ties in Russian respondents' personal networks (Lonkila 1998).

In their comparison of migrant and native St. Petersburg factory workers, Lonkila and Salmi (2005) corroborated the importance of work-related social relations and social support, first to Russian workers in general and, second, to migrant workers in particular. The article at hand builds on the same Russian data corpus as Lonkila and Salmi, but adds to it both an explicit comparison with similar data collected in the neighbouring capital of Finland and an analysis of the structures of the personal networks in each city.

The findings of co-workers' role in Russia run counter to the stereotypical image both of the Russians giving preference to ties with family and kin and of the Finns as a work-centered people. They are, however, in line with Ruan et al.'s (1997) results in China, which stress – similarly in contrast to the traditional image of the weight of kin relations – the importance of work-related ties as a source of social support for Chinese respondents. A replication of a network survey conducted in Tianjin in 1986 and 1993 showed that despite the fact that workplace ties in respondents' discussion networks had been reduced in seven years, their reduction was relatively small in comparison

with the reduction in kin-based ties. Ruan and her associates conclude that the ties with colleagues still played an important role in 1993. Though Lai's (2001) study in a more modern setting in Shanghai partly contested these results, a further comparison between socialist Beijing and capitalist Hong Kong found that the residents of the former were more likely to turn to their co-workers for support then their counterparts in Hong Kong (Lee et al. 2005).

In sum, despite the huge changes at the workplaces with the advance of market relations in China and the fall of the socialist system in Russia (e.g., Ashwin, 1999a,b; Clarke et al., 1996, 1999), the socialist past still seems to be visible in the role of co-worker in support networks both in China and Russia. The remaining text will focus on the comparative analysis of the Russian support networks.

3. Study Sites

This article investigates the social support networks of workers in two different but nationally equally important Russian and Finnish workplaces. The Kirov plant in St. Petersburg was a crown jewel of Soviet factories, employing around 40,000 workers in its heyday and producing tanks, turbines and other machinery. The fall of the Soviet Union forced the factory to reorganize its ownership during the process of privatization and to adjust to the demands of the emerging Russian market economy. By the time of our data collection in Russia in winter 2000, the number of employees had been cut to less than a quarter of the Soviet-era figures. (Lonkila and Salmi, 2005).

The Finnish data was collected in a Helsinki shipyard during the winter of 2003. The shipyard is an integral part of the history of the Finnish shipbuilding industry, boosted after the Second World War by war reparations to the Soviet Union. The Finnish-owned industry did not survive the tough competition, despite a merger in the late 1980s, and the shipyard was bought by a giant Norwegian enterprise in 1991. The early 2000s were marked by layoffs, the

The Importance of Work-Related Social Ties in Post-Soviet Russia: The Role of Co-workers in the
Personal Support Networks in St. Petersburg and Helsinki

CONNECTIONS

number of Finnish employees being cut from roughly 4500 in 2001 to 3600 in 2004.

Both plants had both real and symbolic significance for their home cities. Not only was the Kirov plant named after the Leningrad party leader, but the factory premises cover an immense area in the Kirov city district (*Kirovskii raion*) in St. Petersburg and the Kirov workers, *kirovtsy*, earned a national reputation as exemplary workers of the Soviet empire (for studies of the Kirov factory, see Miroschnichenko and Maksimov, 1994; Grant, 1999). The Helsinki shipyard is similarly a visible part of the city center where immense cruise ships were built until early 2004. The Helsinki shipyard was also well-known nation-wide but, unlike Kirov, this was because of repeated industrial disputes and strikes, particularly during the 1970s.

Both factories were struggling to survive in the globalizing markets and the reorganizations and layoffs kept workers in both cities in a constant state of insecurity. At the time of the collection of the Finnish data, the respondents had already been apprised of forthcoming dismissals and many questioned the future of the whole Helsinki shipyard. In summer 2005, the Norwegian mother company announced it would move its head office from Helsinki to the city of Turku on the south-western coast of Finland. The big cruising ships would be built in Turku and the Helsinki shipyard would focus on smaller vessels, repairs and research. However, a South Korean shipbuilding giant bought the majority of the company shares in 2008 and the speculations about the future of the Helsinki shipyard continued in 2009.

METHODS

This section draws from the description by Lonkila and Salmi (2005) who analyze in detail the differences between native and migrant Kirov workers. For a more detailed description of the data collection and questionnaires used, see Lonkila and Piipponen (2002).

The St. Petersburg interviews took place in one department of the Kirov factory. A total of 50 workers, of whom 12 were women, were interviewed, and their personal networks contained altogether 711 members. The Helsinki data consisted of interviews with 19 male workers, whose networks contained 190 network members. In order to preserve comparability, only male Russian workers were selected and two elderly male workers (69 and 71 years) were excluded. This resulted in the complete data corpus consisting of 36 Russian and 19 Finnish respondents, and of their 490 and 190 personal network members.

The personal networks were constructed with the help of name generators adapted from Claude Fischer's network study *To Dwell Among Friends. Personal networks in Town and City* (1982, cf. Grossetti, 2007, Fischer and Shavit, 1995). These name generators covered several daily-life situations such as with whom the egos talk about work matters (ng1), whose opinion they would listen to when making an important decision (ng2), with whom they shared a common hobby (ng3) or spent free time (ng4), to whom they would turn for such help as repairing domestic appliances or fixing a car (ng5), for baby-sitting or borrowing kitchen utensils (ng6), from whom they could ask to borrow a large sum of money (ng7), to or from whom they had given or received favours during the last three years (ng8), and with whom they had participated in meetings, demonstrations, gatherings or strikes during the last three years (ng9). Finally, the respondents were asked whether there were any important people who had not been mentioned (ng10).

For each name generator, the respondent could name (by first name and initials or by an invented code name) as many people as he wanted. The list of all names given – complemented by the respondent's household members – constitutes the personal network of the respondent. The respondent was then asked to record information about each network member such as age, occupation, place of birth and residence, type and duration of relationship

CONNECTIONS

The Importance of Work-Related Social Ties in Post-Soviet Russia: The Role of Co-workers in the
Personal Support Networks in St. Petersburg and Helsinki

between respondent and network member and how they got acquainted, in a structured questionnaire.

In addition to the questions concerning the personal network, the questionnaire requested basic socio-economic information about the respondent, as well as information about his participation in social and political activities. Moreover, a thematic interview was conducted with each respondent to construct an account of his life course and important life events. Finally, an N x N matrix of each respondent's network members was constructed by asking the respondent to indicate which of the network members had been in mutual contact.

Four methodological points of the study are worth emphasizing. First, the study employs a strictly defined notion of personal network which allows investigation of the totality of the respondents' daily social relations (including friends and relatives, for example). In contrast to confining the study to the work sphere, only this approach enables analysis of the differences in the mixing of professional and personal spheres of life (cf. Gribaudi, 1998; Eve, 2002; Lonkila, 1999). Second, instead of examining values or attitudes toward work, the focus is on the actual micro-level interaction practices. Third, the study joins those students of post-Soviet Russia who stress the importance of investigating social processes at the grass-roots level (e.g., Burawoy and Verdery, 1999, Ashwin, 1999a). Finally, the study is explicitly comparative.

Because of the non-representative sample, the study does not aim at generalizable results. Rather, it seeks to demonstrate the potential of the micro perspective and network methods in comparative studies and to generate fruitful hypotheses for further research.

RESULTS

The average size of the Russian networks in the data corpus was significantly larger than that of the Finnish ones, with 13.6 (SD=3.0) network members in St. Petersburg as opposed to 10.0

(SD=4.8) in Helsinki (p=0.001 in t-test). Not unexpectedly, a majority of the network members in both cities were men, but neither the proportion of male network members in St. Petersburg (67%) nor the mean age of network members (43.7 years) was significantly different than in Helsinki (61% and 43.2 years respectively).

In the following text, the importance of co-workers in each city will be studied using four indicators concerning the personal networks (cf. Piipponen, 2004, Lonkila and Salmi, 2005). These indicators include:

- proportion of co-workers in the networks relative to the number of all personal network members
- overlap (multiplicity) of the various types of informal support and forms of social interactio
- proportion of co-workers who were simultaneously considered as friend
- number of links connecting co-workers with other network members

The first indicator of the co-workers' role is their number in the networks. Because of the difference in the size of the networks in the two cities, this number was calculated relative to the total size of the network. The results showed that the average proportion of colleagues in the St. Petersburg data was more than twice as high (M=33.5%, SD=16.4) than in Helsinki (M=15.4%, SD=12.3, p<0.00). Since the networks were constructed in this study by adding the respondents' household members to the list of people recorded through the ten name generators, the proportion of co-workers in the networks shown above is in itself also an estimate of their importance in terms of mutual support in the workers' lives. Moreover, of the 36 Russian networks studied, 35 (97%) contained at least one co-worker, while the corresponding figure for the 19 Finnish networks was 15 (79%).

Second, the relations between the Russian respondents and their co-worker-alters were

The Importance of Work-Related Social Ties in Post-Soviet Russia: The Role of Co-workers in the
Personal Support Networks in St. Petersburg and Helsinki

CONNECTIONS

more multiplex than those in Finland: The average number of name generators, in which the co-worker-alters were recorded, was significantly greater in St. Petersburg (M=1.56, SD=0.77) than in Helsinki (M=1.13, SD=0.35, p=0.045).

Third, while 97% of Russian respondents and 79% of the Finns reported at least one friend in their networks, only 16% of the Finns reported at least one friend who was simultaneously a co-worker, whereas 64% of the Russians did (respondent could record one network member simultaneously as a friend, co-worker and neighbor, for example).

In sum, these observations speak of the co-workers' significant role as sources of support, and of the blurring of professional and personal spheres of life in Russia. The first three indicators show that the co-workers were relatively more numerous in the networks of Russian workers compared to the Finns; that the Russian workers' relations with co-workers were more varied or multiplex; and that more Russian respondents had co-worker friends in their networks than the Finns. These results reinforce the impression of the significance of co-workers in post-Soviet Russian society vis-à-vis Finnish society.

In the remaining part of this section *the fourth aspect,* namely the structural significance of co-workers in St. Petersburg and Helsinki, will be investigated. The data will be limited to the 35 Russian and 15 Finnish networks containing at least one co-worker. The section is based on the examination of the N x N matrixes of the interconnections between alters filled in by Russian and Finnish respondents. For each alter in the network, the respondent was asked if s/he had been in mutual contact with other alters. The resulting binary matrixes were analyzed with the UCINET network analysis software (Borgatti et al., 2002). A comparison of the basic indicators on the networks in the limited data showed a significant difference in size (M=13.7, SD=3.0 in St. Petersburg, M=10.5, SD=4.9 in Helsinki, p=0.008) and average distance (M=1.50, SD=0.26 in St. Petersburg, M=1.32, SD=0.28 in Helsinki), but no significant difference in density (57% vs 61%) or compactness (0.74 vs 0.71).

The structural significance of co-workers stresses the fact that their role in the networks cannot simply be measured by their number, because any number of co-workers may be weakly connected to the rest of the network. This is exemplified by the following network graph from the Finnish data (Figure 1 – note that the ego is not shown in the figure):
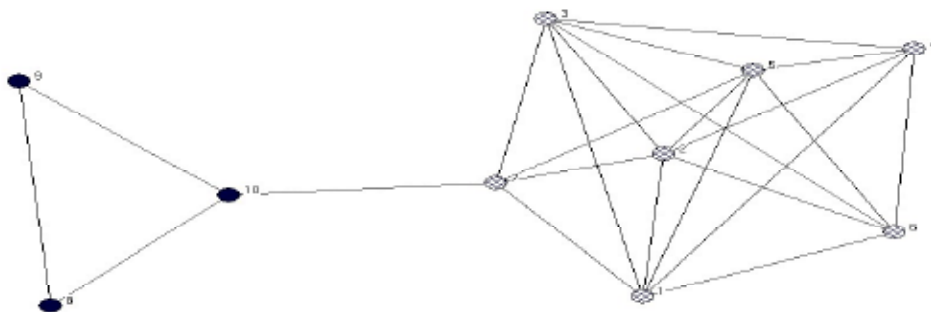
Figure 1. Weakly Connected Clique

Figure 1. Example of a weakly connected clique of three co-workers (black nodes) in the network of a Helsinki worker (hki06).

CONNECTIONS

The Importance of Work-Related Social Ties in Post-Soviet Russia: The Role of Co-workers in the
Personal Support Networks in St. Petersburg and Helsinki

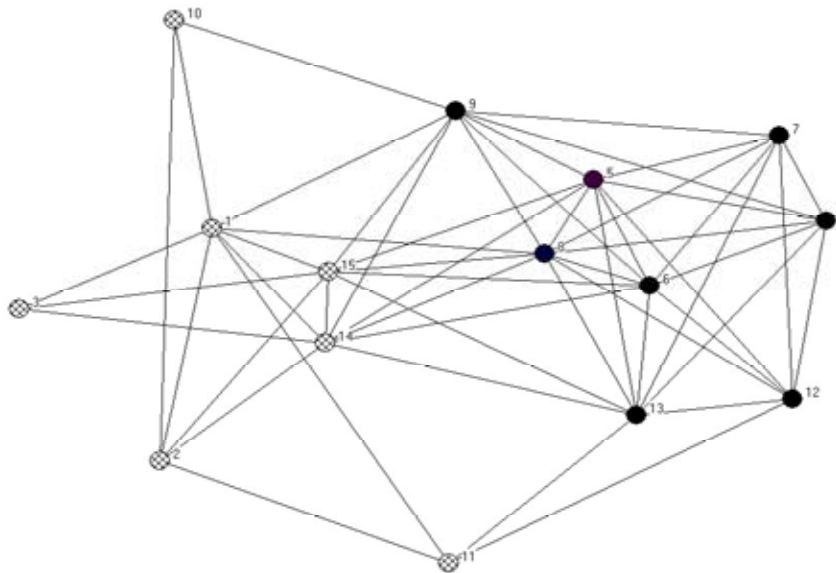**Figure 2. Tightly Connected Co-Workers**

Figure 2. Example of tightly connected co-workers (black nodes) in the network of a St. Petersburg worker (spb 401).

In Figure 1, the connection between co-workers and other network members may vanish – except for the ties with ego – by cutting the "bridge" between the clique of three co-workers on the left-hand side and the rest of the network consisting of family and kin and a plumber friend (no. 7). Hence, in this type of network the co-worker may disappear from the total network without doing much damage to the structure of interaction among the remaining network members. Figure 2 gives a contrary example of a Russian network where the co-workers are much more strongly connected to the whole network structure.

In addition to the mere number of co-workers in the networks, their integration was therefore measured as the number of links connecting egos' co-workers to the remainder of the network. In Figure 1, for example, the total number of links connecting the sphere of work and the rest of the network is one whereas in Figure 2 the corresponding number is fifteen. Based on this reasoning, an indicator of 'co-workers' integration' in the networks was

constructed. Table 1 shows the distribution of this indicator in St. Petersburg and Helsinki.

Table 1 gives additional credence to the general image of relatively low co-workers' integration in Finnish networks as opposed to Russian networks. Eighty percent of the 15 networks in Helsinki but 54% of the 35 networks in St. Petersburg contained less than a quarter of all possible links between the co-workers and other (not work-related) network members.

Though the difference observed in the comparison of the means of this indicator in Russia and Finland was not statistically significant, this was due to one Finnish case only: a network of eight alters, in which the only co-worker in the network knew everyone else of the remaining alters. This raised the integration indicator of this particular case to 100%, not observed elsewhere in the data. After exclusion of this case, the difference between Russian data (M=32.7%, SD=22.7) and Finnish data (M=16.2%, SD=22.3) was significant (p=0.025).

The Importance of Work-Related Social Ties in Post-Soviet Russia: The Role of Co-workers in the
Personal Support Networks in St. Petersburg and Helsinki

CONNECTIONS

Table 1. Indicator of Co-Workers' Integration in the St. Petersburg and Helsinki Networks

	St. Petersburg (N=35)		Helsinki (N=15)	
*Integration indicator**	*no. of networks*	*%*	*no. of networks*	*%*
0-25%	19	54	12	80
25-50%	8	23	0	0
50-75%	7	20	2	13
75-100%	1	3	1	7
Total	35	100	15	100

*The indicator was calculated by dividing the number of the actually effectuated links between co-workers and other network members by the theoretically possible maximum number of these links. The percentages were calculated only for networks containing at least one co-worker.

Lastly, the networks were analyzed with the KeyPlayer 1.1 software programme (*http://www.analytictech.com/*) in order to find three 'key player' nodes for each network, that is, nodes that when removed would result in the largest number of disconnected components. This experiment revealed that in 69% of the Russian networks the three key players contained at least one co-worker whereas the corresponding proportion in the Finnish data was 27%.

There are, however, at least two variables which could easily explain the observed differences (cf. Lonkila and Salmi, 2005). First, the number of co-workers mentioned in the networks is very likely related to the duration of the ego's employment at the factory. The newcomers are generally expected to mention fewer co-workers than those with a long history at the same workplace. Second, the migrants from elsewhere are less likely to introduce their fellow workers to their family and kin living in another part of the country.

The size and nature of our case study data allows only limited control of these variables. A repeated comparison was carried out between "old" migrant workers in the two cities – that is, between the Russian and Finnish workers who

had been working at the factory more than 3.5 years and were thus supposed to have had a chance to get to know their co-workers (cf. Lonkila and Salmi, 2005). The percentage of co-workers, the number of co-workers from whom respondents had received (at least two different kinds of) multiplex support, and the percentage of structural integration of co-workers into the network, were calculated for the 25 Russian and 7 Finnish workers who met these criteria. The results showed that the observed differences either remained the same (percentage of co-workers) or increased (multiplexity and structural integration).

DISCUSSION

The findings of this study are in line with the results of the studies by Ruan and her associates (1997, see also Lee et al. 2005) of Chinese society, suggesting that a socialist system may have effects on networks which outlive its fall and may be resistant to the advance of a market economy. Our findings revealed that Russian co-workers were important as a source of social support in many respects. The St. Petersburg workers' networks not only contained more co-workers than in Helsinki but the ties between the Russian workers and their co-workers were more

CONNECTIONS

The Importance of Work-Related Social Ties in Post-Soviet Russia: The Role of Co-workers in the Personal Support Networks in St. Petersburg and Helsinki

multiplex. In addition, the Russians had more co-worker friends than their Finnish counterparts, and the co-workers seemed to be more densely tied to their personal networks, though the proof of the structural connection remains mixed. Finally, the supportive role of co-workers seems to extend outside the factory walls, thereby blurring the borders of professional and personal spheres of life.

How could these observed differences be explained? Ruan and her associates (1997) related their findings to the continuing legacy of the role of workplace, which controlled most aspects of daily life in communist China:

> "Besides salary, a Chinese workplace typically provided its workers with goods, services, and other material and social advantages such as medical care, housing, loans, child care, and pensions. Many of these benefits, including housing, schools, and services, extended also to the workers' families. It was true also that the distribution of goods and services by the workplace was usually under the control of workshop leaders and other officials. In short, not only did Chinese workers depend on their workplace to satisfy their needs, they depended specifically on influential people at work to obtain needed goods and services." (Ruan et al. 1997, 84)

A thorough understanding of the present-day structure of the respondents' personal networks would require a detailed analysis of their life courses in the specific historical and national contexts. Such a detailed life course study is not attempted here (cf. Lonkila and Salmi 2005). Suffice it to say that, similarly to the description of Ruan et al. (1997) above, the impact of the factory and workplace on the workers' lives was generally much more marked in the Soviet Union than in Finland. The Soviet factory allocated workers jobs and housing, medical care, cars and other goods in short supply, offered them cultural recreation, places to meet other people and so on. Even though many of these benefits were also

provided by Finnish employers, the Finns could also search for solutions to their daily problems on the market (e.g., for housing, cars and other goods), and their social lives were generally much less dependent on the factory than those of their Soviet counterparts.

Explaining the role of co-workers in the present-day Russian networks only as 'Soviet legacy' would be, however, a premature conclusion. The anthropological students of Russian transition have remarked that the features which at first glance look as the remains from the Soviet era, such as the role of barter in Russian economy in the 1990s, may in fact have resulted from the factors and causes unleashed by the transition process itself (Burawoy and Verdery, 1999).

In line with this thinking, the role of co-workers would rather appear as a combined result of the Soviet traditions and post-Soviet experiences: Much of the benefit allocation through the Russian factory has diminished since the fall of the Soviet Union. Having lost the stability and predictability of Soviet era employment, and lacking the unemployment benefits of Finnish workers, post-Soviet Russian workers were more prone to turn to their personal social safety nets, of which co-workers traditionally formed an important part.

Nevertheless, the economic aspects alone can hardly explain the observed differences. Rather, they are more likely to be caused by the complex interaction of social, economic, historical, structural and cultural factors, all of which cannot be addressed in this article.

Our comparison suggests that unemployment or retirement may have different consequences for Russians and Finns. It seems possible that after a working career Finns are more likely to lose the relatively few contacts with their fellow workers than Russians. However, if the Russians also lost these ties, the impact on their social life would be much more grave. In Russia (and probably other post-socialist countries) the maintenance or dissolution of the largely work-related support

The Importance of Work-Related Social Ties in Post-Soviet Russia: The Role of Co-workers in the Personal Support Networks in St. Petersburg and Helsinki

CONNECTIONS

networks may prove vital in the absence of a well functioning social security system.

This study also proposes that the multiplex social ties revolving around the workplace might play a different role in the formation of Russian civil society as compared to western models. In a society penetrated by mistrust in most social institutions and lacking clear interest articulation based on social groups with distinct identities, the workplace-based social networks may function as one possible platform for joint action (Alapuro, 2008; Alapuro and Lonkila, 2000; Gordon, 1997).

Finally, the mixing of professional and personal spheres of life implies that post-Soviet Russian society might combine modern features, such as industrialization and urbanization, with "premodern" aspects, such as weakly differentiated spheres of life and a particularly strong role of networks in economy and society (Srubar 1991, Lonkila, forthcoming). This combination may be indicative of the specific nature of the emerging new socio-political system in Russia and certainly merits further studies.

REFERENCES

Adams, A., Madhavan, S. and Simon, D., 2006. Measuring social networks cross-culturally. Social Networks 28 (4), 363–376.

Alapuro, R., 2008. Kansalaistoiminnan edellytyksiä Venäjällä [Preconditions for civic activism in Russia]. In Salmenniemi, S. and Rotkirch, A. (eds.), Suhteiden Venäjä [Social relations in Russia]. Gaudeamus, Helsinki.

Alapuro, R. and Lonkila, M., 2000. Networks, Identity and (In)action: A Comparison Between Russian and Finnish Teachers. European Societies 2 (1), 65–90.

Ashwin, S., 1999a. Russian Workers: The Anatomy of Patience. Manchester University Press, Manchester.

Ashwin, S., 1999b. Redefining the Collective: Russian Mineworkers in Transition. In: Burawoy, M., Verdery, K. (eds.), Uncertain Transition: Ethnographies of Change in the Postsocialist World. Rowman & Littlefield, Lanham.

Borgatti, S.P., Everett, M.G. and Freeman, L.C., 2002. Ucinet for Windows: Software for Social Network Analysis. Analytic Technologies, Harvard.

Burawoy, M. and Verdery, K., 1999. Introduction. In: Burawoy, M., Verdery, K. (eds.), Uncertain Transition: Ethnographies of Change in the Postsocialist World. Rowman & Littlefield, Lanham.

Clarke, S. (ed.), 1996. Labour Relations in Transition: Wages, Employment and Industrial Conflict in Russia. Edward Elgar, Cheltenham.

Clarke, S.,1999. The Formation of a Labour Market in Russia. Edward Elgar, Cheltenham.

Eve, M., 2002. Les deux traditions d'analyse des réseaux. Réseaux 20 (115), 183–212.

Fischer, C., 1982. What do we mean by `friend'? An inductive study. Social Networks 3 (4), 287–306.

Fischer, C.S., 1982. To Dwell among Friends: Personal Networks in Town and City. University Of Chicago Press.

Fischer, C. S. and Shavit, Y., 1995. National differences in network density: Israel and the United States. Social Networks 17 (2), 129–145.

Gribaudi, M. (ed.), 1998. Espaces, temporalités, stratification: Exercises méthodologiques sur le réseau. Editions de l'Ecole des Hautes Etudes en Sciences Sociales, Paris.

Gordon, L., 1997. The Labour Movement in a Transitional Society: A Force for or against Reform in Russia, International Review of Sociology 7 (2), 337–358.

Grant, J. A., 1999. Big Business in Russia: The Putilov Company in Late Imperial Russia, 1868-1917. University of Pittsburgh Press, Pittsburgh.

Grossetti, M., 2007. Are French networks different? Social Networks 29 (3), 391–404.

Lai, G., 2001. Social support networks in urban Shanghai. Social Networks 23 (1), 73–85.

Lee, R., Ruan, D. and Lai, G., 2005. Social structure and support networks in Beijing and Hong Kong. Social Networks 27 (3), 249–274.

Lonkila, M., 1998. The Social Meaning of Work: Aspects of the Teaching Profession in Post-Soviet Russia. Europe-Asia Studies 50 (4), 699–712.

CONNECTIONS

The Importance of Work-Related Social Ties in Post-Soviet Russia: The Role of Co-workers in the
Personal Support Networks in St. Petersburg and Helsinki

Lonkila, M., 1999a. Social Networks in Post-Soviet Russia: Continuity and Change in the Everyday Life of St. Petersburg Teachers. Kikimora Publications, Helsinki.

Lonkila, M. and Piipponen M., 2002. Social Networks and Civic Culture in Estonia and North-Western Russia. University of Joensuu, Reports of the Karelian Institute 3/2002, Joensuu.

Lonkila, M. and Salmi, A-M., 2005. The Russian work collective and migration. Europe-Asia Studies 57 (5), 681–703.

Lonkila, M., forthcoming. Networks in the Russian Market Economy. Palgrave

Miroshnichenko, N.S. and Maksimov, B.I., 1994. Privatizatsiya Kirovskogo zavoda kak protsess. Sotsiologicheskie issledovaniya 21 (11), 30–38.

Piipponen, M., 2004. Work-related Ties in the Everyday Life of a Russian Karelian Mill Community. In: Alapuro, R., Liikanen, I., and Lonkila, M. (eds.), Beyond Post-Soviet Transition: Micro Perspectives on Challenge and Survival in Russia and Estonia. Kikimora Publications, Helsinki.

Ruan, D., Freeman, L., Dai, X., Pan, Y., and Zhang, W., 1997. On the changing structure of social networks in urban China. Social Networks 19 (1), 75-89.

Srubar, I., 1991. War der Reale Sozialismus Modern? Kölner Zeitschrift für Soziologie und Sozialpsychologie 43 (3), 415–432.

Co-authorship in Italian Workshops on Population Studies:
An Analysis with a Network Approach

Giulia Rivellini
Associate Professor, Università Cattolica del Sacro Cuore
Milan, Italy

Laura Terzera
Associate Professor, Università Milano-Bicocca
Milan, Italy

Interactions, exchanges of ideas and cooperation among scholars are important factors for the advancement of scientific knowledge. Conferences represent one of the most suitable occasions to further scientific interactions, stimulated through the contributions presented either by a single researcher or by a group of authors. Using the books of abstracts from four recent Italian conferences on population studies (Giornate di Studio sulla Popolazione, GSP, 1999, 2001, 2003, and 2005), this research provides an empirical analysis of the collaboration patterns observed among the authors of the papers presented. We followed a social network perspective, in order to find out the determinants of scientific cooperation in the field of Italian demographic studies. The factors playing a major role in determining the actors' relationships seem to be related to gender and to the proximity of universities or conference seats. Although a high number of participants are represented by isolated nodes, the most common way of collaborating is a dyadic relationship. The larger collaborations are due mostly to the presence of a small number of leading authors that manage a large number of papers. Productivity and the popularity of leading authors are attributed to their senior positions in research groups or their technical and statistical skills. It is difficult to measure such aspects with an analysis approach that is different from network analysis.

Authors:

<u>Giulia Rivellini</u>, *Ph.D. is associate professor of Demography and Social Statistics at the Catholic University of Milan. Her research interests are mainly connected to the interaction between individual demographic behaviors and social and territorial space. Recently she also started studying topics of sociology of population studies.*

<u>Laura Terzera</u> *is Associate Professor of Demography at the Faculty of Statistics of the University Milan Bicocca Her main research field is international migration. On behalf of Ismu Foundation she coordinates the Lombardy Sampling Survey on foreign population. She is author of books, essays, articles and research reports.*

Correspondence: *Contact Giulia Rivellini at the Dipartimento di Scienze statistiche, Largo Gemelli, 1, 20123 Milano, ITALY. Tel: 0039 02 72343684; fax: 0039 02 72342671; email: giulia.rivellini@unicatt.it.*

INTRODUCTION

Some of the main factors involved in the advancement of scientific knowledge are the interactions, exchanges, and cooperation among scholars, particularly when these interactions occur across disciplines. The analysis of scientific production is useful in understanding both the extent and the evolution over time of topics and methods, as well as in better understanding co-authorship (collaboration) among scientists (Farahat 2002; Liang et al. 2002; Newman 2003). To formulate this kind of analysis, several types of data sources can be considered: 1) electronic bibliographies that index a thematic literature comprehensive of journal articles, books, book reviews, collective volume articles, working papers and dissertations (see for instance Econlit); 2) reviews of thematic references; 3) catalogues of electronic journals; 4) books of abstracts distributed at formal meetings; 5) information systems for the management of government funds for scientific research. These data sources are neither completely overlapping nor interchangeable, nor are they exhaustive, consequently they do not allow for the creation of a comprehensive and unique data base with regards to individual scientific production. As a result, in order to carry out quantitative analyses on scientific production, one is obliged to choose from among the information bases available. In this paper we will consider the fourth data source.

In the demographic field, recent studies following a primarily qualitative approach have concentrated on the themes, methods, and theories used in scientific publications (Tabutin 2005; Hoem 2007). Studies following a quantitative approach show some weaknesses. For example, the article by Gendrau and Huix-Adamets (2003) only takes into account some national research units. Chasteland's paper (2004) analyzes data gathered through an international web survey based on voluntary answers provided by demographers thus creating a sample that is not statistically relevant. In Italy, quantitative analyses are based on books

of abstracts from scientific conventions (Rivellini and Terzera, 2008) or demographers' references collections (Casacchia and Mancini 1995; Rivellini and Rizzi 2002; Rivellini et al. 2006).

There is clearly a lack of quantitative analysis on scientific collaboration primarily due to the difficulty in obtaining an exhaustive database, as noted above. Moreover, there is limited knowledge regarding analysis methods suitable for studying research collaboration. Network analysis is a suitable way to study how the scientific community organizes its relationships (Wasserman and Faust 1994).

Following this approach, the aim of this research is to provide an empirical analysis of the scientific relationships observed among authors of the papers presented in the national workshops on population studies (Giornate di Studio sulla Popolazione, GSP) organized every two years by the AISP (Association of Italian Population Studies, AISP http://gcd.stat.unipd.it/new/about/11/home). This convention attracts the largest number of Italian demographers.

This paper is organized as follows: the first section describes the data-set used in the current analysis. The second section presents the methods and results of a cross-sectional and network analysis of scientific production (with an analysis of one complex network configuration). Thirdly, the case of academics interested in population studies is analyzed. The final section concludes with a discussion of the findings.

METHODS

1. Abstract Collection and Network's Definition

We used all the abstracts collected in the books of abstracts from the four GSP conferences (1999, 2001, 2003, 2005) to build our dataset as demonstrated in Figure 1. In the first part of the figure we find a typical abstract from which we

took information on the paper (title, topic, sources and methods of analysis) and on the author/s (gender, affiliation). In the second part of the figure the information is transferred into the dataset where we introduce a double code for each statistical unit: one corresponding to the presented paper and another to the author, who could have contributed multiple papers to the conference (in this case the variable "Number of contributions" is higher than 1). This procedure allows us to analyze data from the viewpoint of the contribution and collaboration of the author.

The dataset was constructed for two distinct objectives: (1) to describe the themes and methods used in the four GSP conferences (Rivellini and Terzera, 2008), and (2) to outline the existing forms of scientific collaboration between researchers. In this paper we focus on the second aim, jointly analyzing authors and their research contributions. More specifically, we are interested in understanding if Italian demographers generally write a paper on their own or in a group and in finding out how large and homogeneous the groups are in the latter case. For this aim we need to adopt the most common definition of a complete social network, made up of i) a finite group of actor nodes and ii) the relationship (or relationships) linking them together.

Figure 1. From the Abstract to the Dataset: An example

Abstract

> *Title*: **A correlated frailty model with long-term survivors for estimating the heritability of breast cancer**
>
> Isabella Locatelli (*Author 1*)
> *Department of Quantitative Methods, University Luigi Bocconi, Milan, Italy*
> Alessandro Rosina (*Author 2*)
> *Institute of Population and Geographical Studies, Catholic University, Milan, Italy*
> Paul Lichtenstein (*Author 3*)
> *Karolinska Institutet, Stockholm, Sweden*
> Anatoli I. Yashin (*Author 4*)
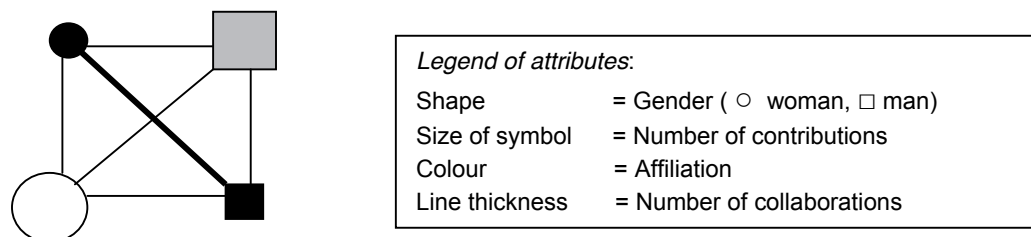> *Max Planck Institute for Demographic Research, Rostock*
>
> **Abstract**
> The aim of this study is to investigate the role of genetics and environment in susceptibility to breast cancer (frailty). An interdisciplinary approach was adopted, combining a correlated frailty-mixture model with genetic equations ....

Dataset

Code_paper	Code_author	Year	Title	Surname Name	Topics	Gender	Affiliation	Number of contributions
403	206	2005	A correlated frailty model	Locatelli Isabella	34 = Methods	0	2 = Academic background	1
403	327	2005	A correlated frailty model	Rosina Alessandro	34 = Methods	1	1=Academic	5
403	482	2005	A correlated frailty model	Lichtenstein Paul	34 = Methods	1	8=Non-university foreigner	1
403	413	2005	A correlated frailty model	Yashin Anatoli	34 = Methods	1	8=Non-university foreigner	1

Figure 2. Example of Co-authorship Network Relative to a Generic Contribution

Legend of attributes:

Shape	= Gender (○ woman, □ man)
Size of symbol	= Number of contributions
Colour	= Affiliation
Line thickness	= Number of collaborations

The actors are usually "social" units (i.e. people, organisations, communities, nations, regions) with the links (i.e. friendship, collaboration, flow of goods, resources, monetary transfers) between them forming a social network. Assuming the network is composed of n actors, it can be represented by a matrix Y (adjacency matrix) with dimensions n x n in which the generic element Y_{ij} supplies the information concerning the relationship between actor i and actor j. The relationship variable pair (Y_{ij}, Y_{ji}) is called a dyad. The matrix is symmetric if the relationships between the nodes are not oriented, and asymmetric when the relationships are oriented.

More specifically in the present case, the generic tie is defined as Y_{ij}, where Y_{ij} =1 if at least one paper is co-authored by scientists i and j (i, j = 1, …, n; $j \neq i$) and Y_{ij} = 0 if otherwise. The characteristics of each author (gender, affiliation, number of contributions and number of collaborations with same actor) are managed as node attributes. The author's potential affiliations are recognized as follows:

- Italian academics, such as full or associate professors, researchers, or *academic entourage* (people who collaborate with the university in a not yet permanent position, for instance PhD and Post Doc students)
- national research or statistic institutes (Istat, Irpps, Iss)
- local public agencies
- foreign universities or research institutes (INED, etc.)

The network perspective also allows for a graphic representation of collaborations in which it is possible to observe relationships among authors as well as the nodes' characteristics, as illustrated in Figure 2.

2. Scientific Co-authorship According to a Cross-sectional and Network Perspective

The nodes are authors, while the relationship that links them together is their collaboration in writing and/or preparing a presentation. The summarized overview of the major network measures observed in the four conferences are presented in Table 1. There is an increase in participation between 1999 and 2003 in both the number of authors, and the number of contributions. In particular, from 1999 to 2001, there was an increase of 122% in the number of contributions and of 96% in the number of authors, while from 2001 to 2003 there were 28% more contributions and 24% more authors. This reveals an increasing number of demographers and other researchers in the field of population studies.

In the 2005 edition, the organizers, given the increased memberships in preceding years, redefined the rules to participate in the GSP[1] introducing a limited number of contributions and thus stricter selection criteria for papers. They also introduced a poster session that gathered approximately 14% of the contribution set. This "forced" reorganization caused a strong

---

[1] Until 2003, the spirit behind the GSP can be concisely defined as a confrontation/debate on presented works even if not fully completed and therefore without any preliminary review process.

decrease in participation in the final year considered by this study.

Another interesting element to highlight is that the success of the GSP is due in great part to the high level of turnover of the participants; 74% of the authors had taken part in only one year, 17% twice, and a very small percentage of researchers (3%) was present in all four of the years considered (see Rivellini and Terzera, 2008).

Table 1. Network Measures by Year

	1999	**2001**	**2003**	**2005**
No. of Contributions	54	120	154	110
No. of Authors	102	200	248	180
Density	0.220	0.014	0.009	0.012
Inclusivity	0.882	0.815	0.859	0.811
Average Degree	2.176	2.800	2.290	2.160
Variance of Degree	3.691	12.69	3.5	3.85

The visualization of the collaboration network observed cross-sectionally for each of the four years, illustrates the evolution over time of collaboration methods, and the consistently low density of the four networks (Figure 3). The number of contacts between the various individuals is, in fact, never higher than one fifth of 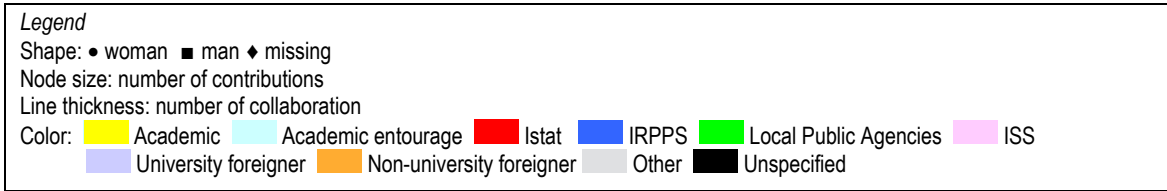all possible contacts. The density overall is quite low and mostly decreases between 1999 and 2003. This can be explained by both the decrease in the number of lines in relation to the number of nodes, and the progressive, though slight, increase in the number of isolated nodes illustrated by the reduction of the inclusiveness measure (the ratio of non-isolated nodes is always higher than 80%, with maximum value in 1999 and minimum value in 2005).

Concentrating on the various forms of collaboration that emerge from the four graphs, a prevalence of small groups of researchers can be noted. These are clustered in different sub-graphs, except for the component 1-node and 2-lines connected in the 2001 network that involves a large number of authors (n = 35) with the same affiliation (local public agencies). Later we will better illustrate some additional specificities of this sub-graph.
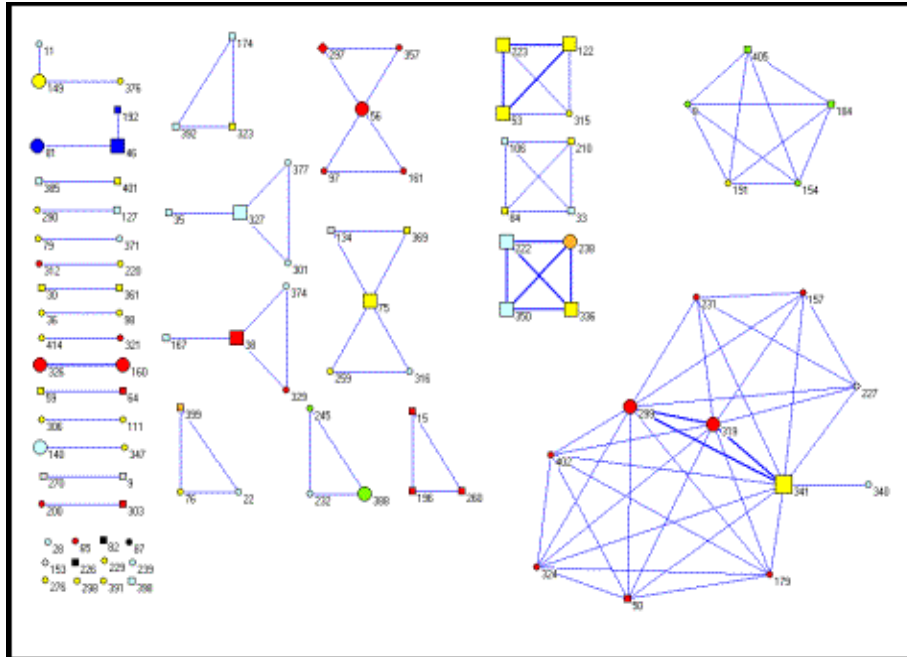
We can therefore note how the network structure generally shows a pattern of low cohesion, because small components are recognized with a majority represented by dyadic relationships. This is true for every year analyzed, with the exception of 2001, which was characterized by an average degree equal to 2.8. This appears as a quite high value due a large number of contributions co-authored by large groups.

An increase in women's participation can be globally observed, starting in the 1999 edition. Moreover, among women a clear, if steadily decreasing inclination to produce individually can be noted (75% of isolated nodes are women in 1999 decreasing to 51.4% in 2005).

**Figure 3. Collaboration Network by Gender, Affiliation, Number of Contributions,
Number of Collaborations Inside the Same Research Group**

Legend
Shape: • woman ■ man ♦ missing
Node size: number of contributions
Line thickness: number of collaboration
Color: ▢ Academic ▢ Academic entourage ▢ Istat ▢ IRPPS ▢ Local Public Agencies ▢ ISS
       ▢ University foreigner ▢ Non-university foreigner ▢ Other ▢ Unspecified

a) 1999

b) 2001

c) 2003

d) 2005

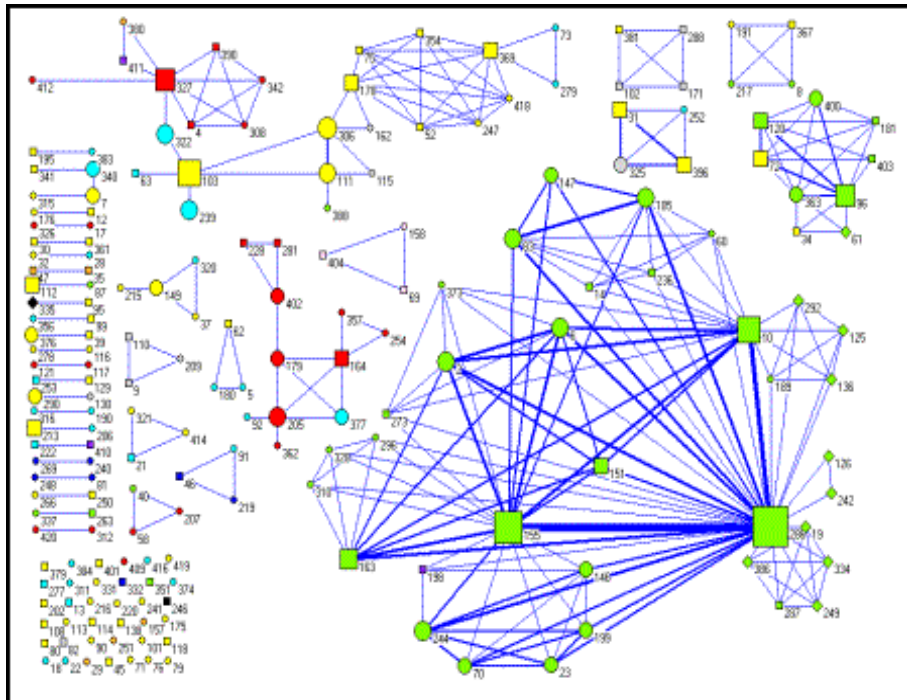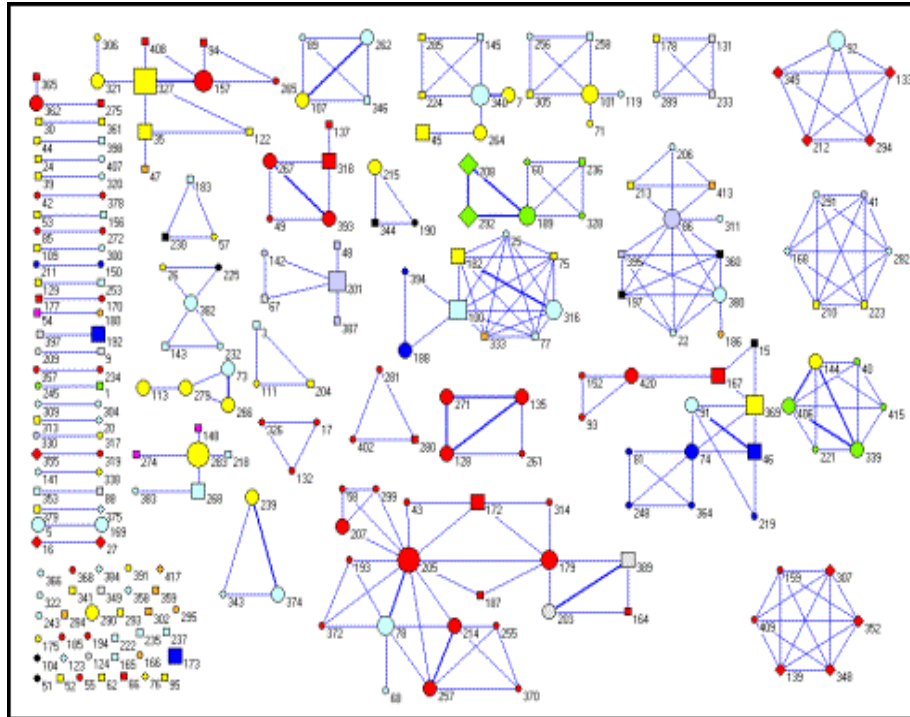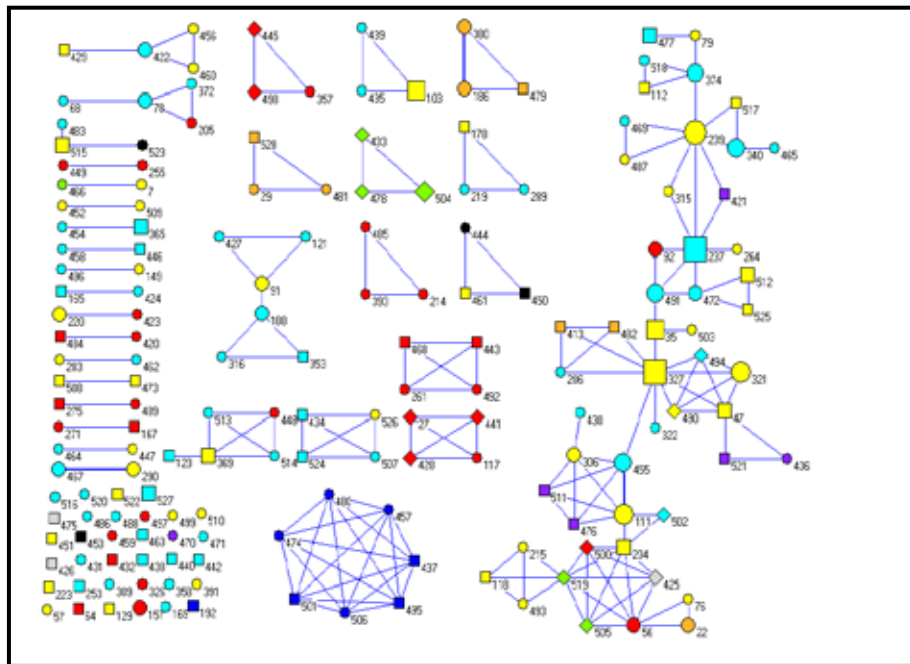**Figure 3a,b,c,d.** Collaboration networks are shown. Gender, affiliation, number of contributions, and number of collaborations inside the same research group are indicated. Software Pajek was used to draw a map of the co-authorship networks.

Production intensity is measured by the number of papers produced by each author. In 1999, there is a balance between genders, while in 2001, 2003, and 2005 women appear more productive, even though the record for most contributions presented at a single conference is held by a man. From Figure 3-b, it can be seen that author productivity might be associated with central position in the network (see the following paragraph). Finally, considering the affiliation (colour of the nodes), we can see that academics and their entourage are among the most assiduous participants in the GSP (Table 2). This can be explained by the necessity of scientific productivity for career development.

Table 2. Authors by Affiliation and Year of Conference (%)

Affiliation	1999	2001	2003	2005
Academic	32.35	32.35	21.37	27.78
Academic Entourage	23.53	15.00	23.79	33.33
Istat	25.49	12.00	27.82	18.89
Irpps	2.94	3.50	4.84	4.44
Local Public Agencies	5.88	24.00	5.24	3.33
University Foreigner	-	1.50	3.23	3.33
Non-university Foreigner	1.97	3.00	5.63	5.00
Other	4.90	7.50	4.44	1.67
Unspecified	2.94	1.00	3.64	2.00
Total	*100.00*	*100.00*	*100.00*	*100.00*

Sorce: Rivellini and Terzera, 2008

A substantial openness over time towards people from various Italian or foreign agencies can be detected. In particular, the increasing presence of foreign authors or Italian authors affiliated with foreign organizations (amounting to 8% of the entire population over the last two years) could also be an indication of the expansion outside of Italy of this particular conference typology, as well as an indication of an increase of transnational collaborations among authors.

Looking at the collaborations and taking into account first gender and then affiliation a few potential patterns emerge: no clear gender majority exists (43% of the network components in 1999 and 64% in 2005 are mixed gender), although, the smaller the collaboration group is, the more likely it is to detect female homogeneity. Dyads demonstrated the most female homogeneity, while components of greater size share a more balanced mix of author gender (Table 3).

Regarding the homogeneity of affiliation, we separated the authors into groups of academics (academic + academic entourage + university foreigner) and non-academics. We observed a similar pattern to that of gender: a prevalence of academic collaboration in the overall networks (~40%), with greater homogeneity in smaller components (especially for academic affiliation). The clustering by affiliation therefore has been increasing since 2003 (Table 4).

In general, a high number of participants are represented by isolated nodes (the most isolated are Irpps - Institute of Research on Population and Social Policies - and non-Italian agencies). Among the non-academic groups, there is strong homogeneity by local public agencies, Irpps, and Istat (National Statistical Institute). Inclusiveness indicators and average degree have further proven that local public agencies and Istat tend to write in larger groups (Table 5).

Table 3. Contributions by Component Type and Gender Homogeneity

Edition	Component Type	Homogeneity of Male	Homogeneity of Female	Mixed	Total
1999	Dyad	4 (31%)	7 (54%)	2 (15%)	13
	Triad	2 (33%)	2 (33%)	2 (33%)	6
	Quadrad	1 (33%)	0	2 (66%)	3
	Other subgroups	0	0	6 (100%)	6
	Total	*7 (25%)*	*9 (32%)*	*12 (43%)*	*28*
2001	Dyad	4 (18%)	9 (41%)	8 (36%)	22 (1 missing)
	Triad	0	1 (17%)	5 (83%)	6
	Quadrad	1 (20%)	1 (20%)	3 (60%)	5
	Other subgroups	0	0	4 (100%)	4
	Total	*5 (14%)*	*11 (30%)*	*20 (54%)*	*37*
2003	Dyad	6 (24%)	8 (32%)	10 (40%)	25 (1 missing)
	Triad	0	2 (33%)	4 (67%)	6
	Quadrad	0	2 (50%)	2 (50%)	4
	Other subgroups	0	1 (7%)	13 (93%)	14
	Total	*6 (12%)*	*13 (27%)*	*29 (59%)*	*49*
2005	Dyad	1 (7%)	8 (53%)	6 (40%)	15
	Triad	0	1 (11%)	7 (77%)	9 (1 missing)
	Quadrad	0	1 (20%)	4 (80%)	5
	Other subgroups	0	0	4 (100%)	4
	Total	*1 (3%)*	*10 (3%)*	*21 (64%)*	*33*

Table 4. Contributions by Component Type and Homogeneity of Affiliations

Edition	Component Type	Homogeneity of Academic	Homogeneity of Not Academic	Mixed	Total
1999	Dyad	7 (54%)	3 (23%)	3 (23%)	13
	Triad	3 (50%)	2 (33%)	1 (17%)	6
	Quadrad	2 (67%)	0	1 (33%)	3
	Other subgroups	1 (17%)	1 (17%)	4 (66%)	6
	Total	*13 (46%)*	*6 (22%)*	*9 (32%)*	*28*
2001	Dyad	12 (55%)	7 (32%)	3 (14%)	22
	Triad	2 (33%)	3 (50%)	1 (17%)	6
	Quadrad	1 (20%)	0	4 (80%)	5
	Other subgroups	0	0	1 (100.0%)	1
	Total	*15 (40%)*	*10 (27%)*	*12 (33%)*	*37*
2003	Dyad	14 (56%)	10 (40%)	1 (4%)	25 (1 missing)
	Triad	2 (33%)	2 (33%)	2 (33%)	6
	Quadrad	2 (50%	1 (25%)	1 (25%)	4
	Other subgroups	3 (21%)	2 (14%)	9 (65%)	14
	Total	*21 (43%)*	*15 (31%)*	*13 (26%)*	*49*
2005	Dyad	9 (60%)	4 (27%)	2 (13%)	15
	Triad	2 (22%)	5 (56%)	2 (22%)	9
	Quadrad	2 (40%)	2 (40%)	1 (20%)	5
	Other subgroups	1 (25%)	1 (25%)	2 (50.0%)	4
	Total	*14 (42%)*	*12 (36%)*	*7 (21%)*	*33*

Table 5. Inclusivity and Average Degree by Affiliation

Affiliation	1999		2001		2003		2005	
	Inclusivity	Average Degree	Inclusivity	Average Degree	Inclusivity	Average Degree	Inclusivity	Average Degree
Academic	0.879	1.91	0.677	1.82	0.830	2.02	0.840	2.30
Academic entourage	0.875	1.63	0.733	1.17	0.831	2.27	0.750	1.72
Istat	0.962	3.15	0.958	2.54	0.928	2.57	0.824	1.88
IRPPS	1.000	1.33	0.857	1.14	0.917	2.33	0.875	5.25
Local public agencies	1.000	3.33	0.979	6.13	1.000	3.15	1.000	3.33
ISS	-	-	1.000	2.00	-	-		
University foreigner	-	-	1.000	2.33	1.000	3.50	0.833	2.33
Non-university foreigner	1.000	2.50	0.500	0.67	0.455	1.09	1.000	2.22
INED	-	-	-	-	1.000	1.00		
Other	0.800	1.80	0.917	2.17	0.818	1.64	0.333	1.67
Unspecified	0.000	0.00	0.500	0.50	0.778	2.22	0.750	1.25

A Deeper analysis on the most complex sub-graph. The presence of a complex sub-graph is apparent in the 2001 network, which allows for a more in-depth analysis regarding scientific collaborations from a network perspective. This sub-graph, recognized as a 2-clique network, is a group with a higher number of green nodes. This is established by one node's ability to activate multiple links with authors of various papers. The sub-graph is denser than other sub-graphs in Figure 2. A density level equal to 0.212 illustrates the presence of more than 21% of the possible links observed with 35 nodes. The potential number of lines is 595 and the average degree is 7.2, a value rather far from the maximum theoretical value of 35. The variance associated to this indicator is quite high (33.224), because of the degree of single nodes that vary from a minimum of 2 to a maximum of 34. This signifies that the authors concerned by this part of the network collaborated in large groups, with at least 3 people, a specificity typical of non-academic researchers. Indeed in this sub-graph are included three vertexes with the highest number of intersecting lines: the node 286 can be cited as the maximum degree,

155 with degree 20, and 10 with degree 17. These three actors are also very collaborative among themselves: 8 papers between nodes 155 and 286 and 5 between nodes 10 and 286.

Bearing in mind the color of the nodes, the prevalence of green demonstrates that all individuals, excluding 198, work for local public agencies, specifically for the "Agenzia di Pubblica Sanità della Regione Lazio" (Health Local Public Agency, Lazio region). The scientific behavior of this group seems particularly dependent on node 286. In fact, if we consider the contributions produced, it can be observed that all are involved with this vertex, which exhibits the largest size. The key role played in the sub-graph by the node 286 is also confirmed by the following remarks regarding the location of actors. In this sense the centrality measures give us an insight into various roles: who are the connectors, who is in the core of the network, and who is on the periphery.

The sub-graph is connected because all possible couple points are accessible through a path.

Node 286 is a 'connector' or 'hub', because it shows the highest number of direct connections. Next to this vertex, actors 155 and 10 play an important role in the network, even if their *degree centralities* are lower compared to that one of the connector (0.588 and 0.500 respectively for node 155 and 10 versus 1 for node 286). Nodes 126 and 242 are on the periphery, with a normalized degree centrality equal to 0.059. However, similar observations regarding each node's roles can be made if we consider betweenness centrality: the nodes with the best locations are still 286, 155, and 10. These nodes play a powerful role as 'brokers' in the sub-graph, but they are also single points of failure at the same time.

The relationship between the centralities of all nodes reveal a very centralized network that is dominated by one or a few very central nodes, with high degree and betweenness centrality. In this case these nodes correspond to three men who produced the highest number of papers co-authored.

These network measures also point out the skills required to be a 'central' researcher: managing a large number of papers, collaborating in a team; being in a leadership position in the research groups or having a technical background useful for doing statistical analyses.

The prestige of single nodes is thus higher when the whole sub-graph is neatly structured around the more central nodes. The importance of one node within the entirety of scientific collaborations is finally confirmed by other network characteristics, associated with the analysis of the sub-graph connection. The concepts linked to this aspect are the following: *cut-point*, *bridge*, *point-connectivity* and *line connectivity*. It can be noted, for example, that node 286 is not only the most central point, but also a *cut-point*, as removing it from the network we would obtain a larger number of *components* compared to the sub-graph that includes it (the same would not happen if nodes 155 and 10 were eliminated, since node 286 has the maximum level, the graph would then again

form a sole *component*). Moreover, excluding node 286 from the network, the new graph would end up disconnected. The result is that the minimum number of nodes that have to be removed in order for the graph to be disconnected is 1 and hence the sub-graph is defined as *1-node connected*.

If we look at the lines, instead at the nodes, the minimum number of the lines that must be removed for the graph to be disconnected is 2 (those that link node 286 to actors 242 and 126). This means that the sub-graph is a *2-lines connected*. This brief description of the sub-graph has emphasized how network characteristics' analysis, which can be empirically obtained through quantitative indicators, allows us to qualify each single author with reference to his/her capacity to interact diffusively with the rest of the scientific community. This particularity seems related to the productivity and the popularity of the researcher, a quality that is difficult to jointly measure with an analysis approach different from network analysis.

3. The Case of the Academics

Attention will now be placed on the sub-group which participates most frequently in the GSP, that is, the group of academics. Among them the demographers play a major, though not exclusive, role. In fact, their presence has a percentage which fluctuates between 54% (edition 1999 and 2005) and 60.4% (2003), leaving ample space for participation by academics of other scientific sectors (Social Statistic, Anthropology, Medicine, Economics) and offering to the GSP a strong interdisciplinary quality.

Next, a detailed analysis focuses on characteristics of the sub-group, followed by the examination of scientific relationships maintained between academic demographers and researchers in other fields. Concerning the first aspect, through the examination of the maps in Figure 4, it is possible to observe the geographic allocation of the participating

academics on the Italian territory, its evolution over the various conferences and, moreover, some of the prominent structural characteristics of the sub-group actors.

In particular, with the succession of conferences, the territorial distribution based on the origin of such participants has seen an increase in its own heterogeneity. Initially, the origins were concentrated in Central Italy and secondly in the North, while the South and the islands were scarcely represented. Over time, the North has reinforced itself, with Milan becoming the most "active" node in 2005, as the presence of Piedmontese academics consolidated itself. In the South, as of 2003, the participation of researchers from the Neapolitan area emerges; but even more noticeable is the area around the Bari region that has registered a constant presence in the GSP since 2001.

The only territorial entities that appear excluded, or at the most rarely present, are the islands. These trends are linked to the different expansion of the academic staff according to their geographical origin. On a national view, in fact, professors of Demography have increased (between 2001 and 2005) by approximately 2% compared to a little over 10% for all non-demographer academics (but coming from a scientific sector represented at least one time in the conferences). Nevertheless, in both cases, the main increase is observed in Southern Italy, respectively from 21 to 24% roughly (Ministry of Education, *www.miur.it*).

A second element that can be observed is the "dependence" on the place in which the GSP were organized. According to the years, it can be observed that an increasing number of authors come from bordering or relatively nearby geographical areas. This trend is most apparent in the case of the last two conferences, where academics from the areas around Bari and Padua, respectively, appear consistently, and which turns out to be a common characteristic

even in more recent international conventions that deal with demographic themes (Rivellini and Terzera, 2008). Again, for all the conferences considered, a limited interaction between groups of different territorial domains can be observed. More frequent, however, are the scientific collaborations between researchers belonging to faculties of the same university or located in the same or nearby cities. With such an assertion, we can observe how scientific interaction is further facilitated and/or conditioned by geographic proximity.

With regards to the participation of various positions (Figure 4), a substantial balance can be seen between full professors, associate professors, and researchers, even if full professors have been more prevalent since 2003. Keeping in mind that this change is also conditioned by the modifications carried out from the reference population between 2001 and 2005. Between academics of the sectors present in the various conferences the increase in full professors is approximately 20% while researchers and associate professors increase only by 6-7%. Such a difference is even more distinct between full professors of Demography who show an increase of 25% compared to a reduction in researchers (of 22%) and a significant freeze in the number of associate professors (*www.miur.it*).

Another element of interest is the relationship between the type of collaboration of academics and their position (Table 6). These figures show that full professors collaborate, more than other academics, with authors belonging to both universities and other organizations. While among associate professors, collaborations with researchers belonging to organizations other than universities are more frequent. And finally, academic researchers tend to present contributions individually or with other academics (researchers, associate and full professors), more commonly than the other two academic groups.

Table 6. Participating Academics (all years) by Status and Collaboration Type (%)

Collaboration Type	Researchers	Associates	Full Professor
Alone	25.8	16.7	19.2
With organisations	21.0	43.3	26.9
With academics	29.0	21.7	16.7
With organisations & academics	24.2	18.3	37.2
Total	100.0	100.0	100.0

Looking more closely at the indicators of network analysis, in 1999 and 2001, the density of the subgroup was lower than the one observed on the whole network: so the academics seem to collaborate more often in small groups or to do research individually. Such indices have increased since the 2003 conference and exceeded the values achieved by the entire population, although overall values remain low (Table 7).

On the other hand, even the inclusivity value indicates how these participants are not, in an exclusive manner, inclined towards collaborations with other academics (no more than 50% of these have some kind of scientific link). Even the *average degree* is always lower than the one registered for each participant and, other than 2001, lower than 1. This shows, once again, how academics prefer collaborations that either strictly involve members of the university or are completely outside this circle.

Regarding the distribution by gender (Figure 4) in the GSP overall, the participation is by and large equally distributed between the two sexes, even though during the various conferences slight oscillations can be observed corresponding to a major presence of women in 1999 and in 2005. For example, in 2005 women professors represented 56% of the participants, while in Italy itself only 37.2% of the whole academic group are female (this percentage rises

to 47% in the case of professors of Demography (*www.miur.it*). But if we also take into account associate and research positions, a female prevalence is always present (overall in the GSP, 59% of both sub-groups are women), while among full professors men dominate in all conferences (62% of all GSP). This is further proof of the larger female participation in these types of conferences given that female percentages are considerably lower in the academic reference population, especially in relation to associate and full professors (38.3% and 21.1% in 2005, respectively).

Table 7. Network Measures of Academics

	1999	2001	2003	2005
No. of Authors	33	65	53	50
Density	0.030	0.002	0.011	0.018
Inclusivity	0.515	0.477	0.358	0.480
Average Degree	0.788	1.077	0.604	0.880
C.V.	1.241	1.718	1.227	1.312

If we observe the collaborations with respect to gender (Figure 4) maintained over the four years, we can see an increase in collaborations which have already emerged through the network indicators: the percentage of those that present papers individually since 2003 is in fact decreasing across genders. Among male academics, this tendency is less apparent: if in the first two years considered, more women presented works on their own, in the last two editions of the GSP it is the male academics that have taken this approach. Again, in the last two years considered, the collaborations between academics are only slightly more common among women, without having a noticeable impact among other types of collaboration. In fact, with regards to collaboration, no consistent predominance of one gender over the other can be observed.

The network indicators have shown how the collaborations among academics are quite limited and restricted. This is made clear in the third graph of Figure 5, which shows a decrease over time and for both genders of these types of relationships and an increase, on the contrary, of more diverse collaborations. As previously observed, demographers are the primary, but not exclusive participants in the GSP. In particular, Figure 6 shows that it is possible to see an increased participation by statisticians over time and a steady - although varying - participation by economists, anthropologists and doctors. However, we are confronted with the limited interaction with sociologists (because of low count labeled "Other"), despite these collaborations' relevance and desirability Stoetzel asserts: **"**… Problems of social action may be an effect of Demography; there are social inadequacies whose causes are demographic. … Population aging will affect not only the economy and employment, but also the arts and politics. It is not an exaggeration to say that Demography largely commands social life and that every sociologist should also be a demographer**"** (Stoetzel 2006, p. 25).

**Figure 4. Geographical Location of Academics by Gender and Status: Florence (1999),
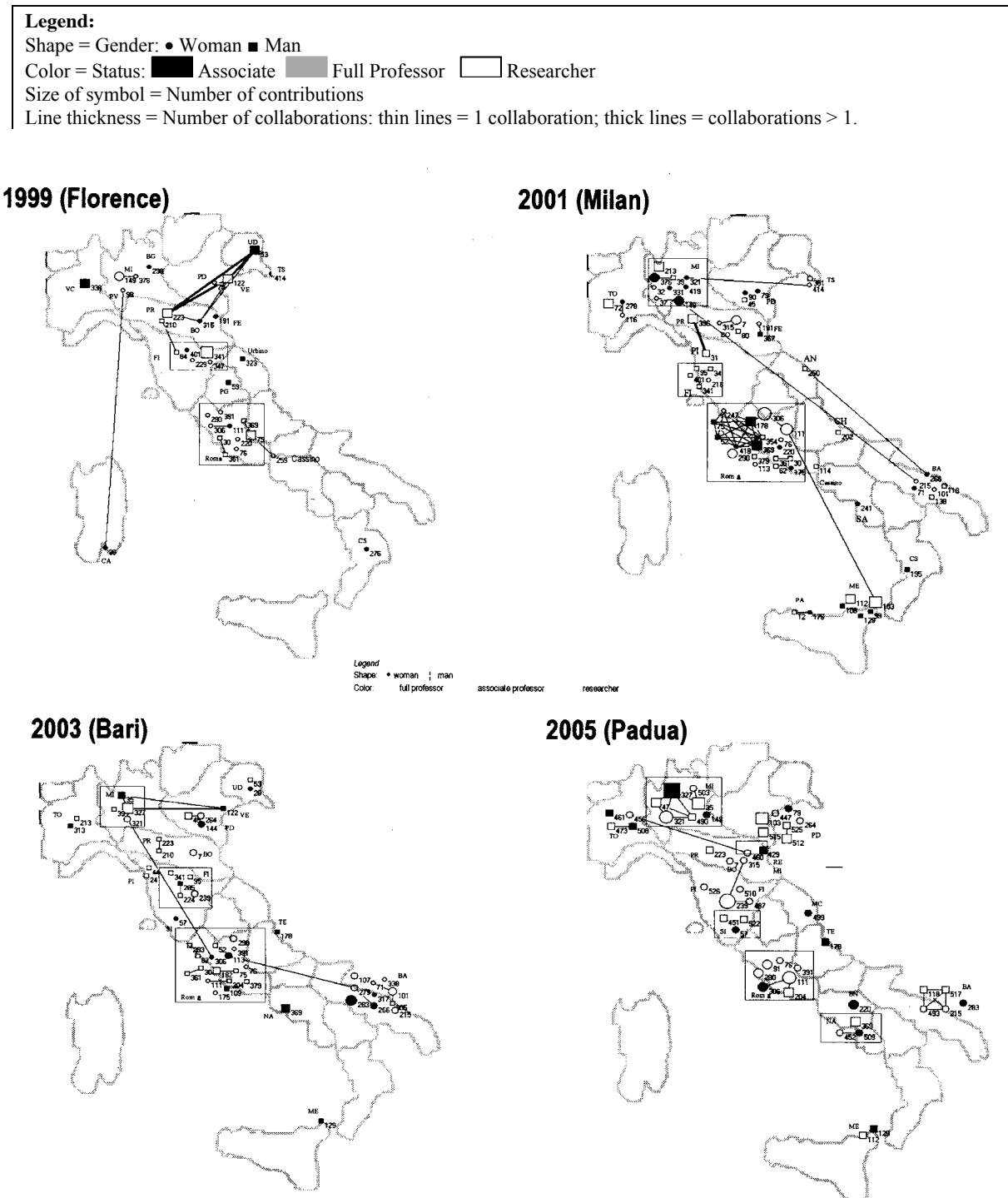Milan (2001), Bari (2003), Padua (2005)**

Legend:
Shape = Gender: ● Woman ■ Man
Color = Status: ■ Associate   ▨ Full Professor   ☐ Researcher
Size of symbol = Number of contributions
Line thickness = Number of collaborations: thin lines = 1 collaboration; thick lines = collaborations > 1.

1999 (Florence)

**2001 (Milan)**

2003 (Bari)

2005 (Padua)

Figure 4. The figure displays for every edition the geographical distribution of the cities in which the universities of the GSP participating academics are located. When we identified many participants coming from the same big cities, such as Florence, Rome or Milan, we highlighted them in a square. To map the points we used the geographical coordinates in sexagesimal degrees available at: http://www.satellitedidattico.it/it/supporto/coordinate.asp.

Figure 5. Academic Authors by Kind of Co-authorship and Gender
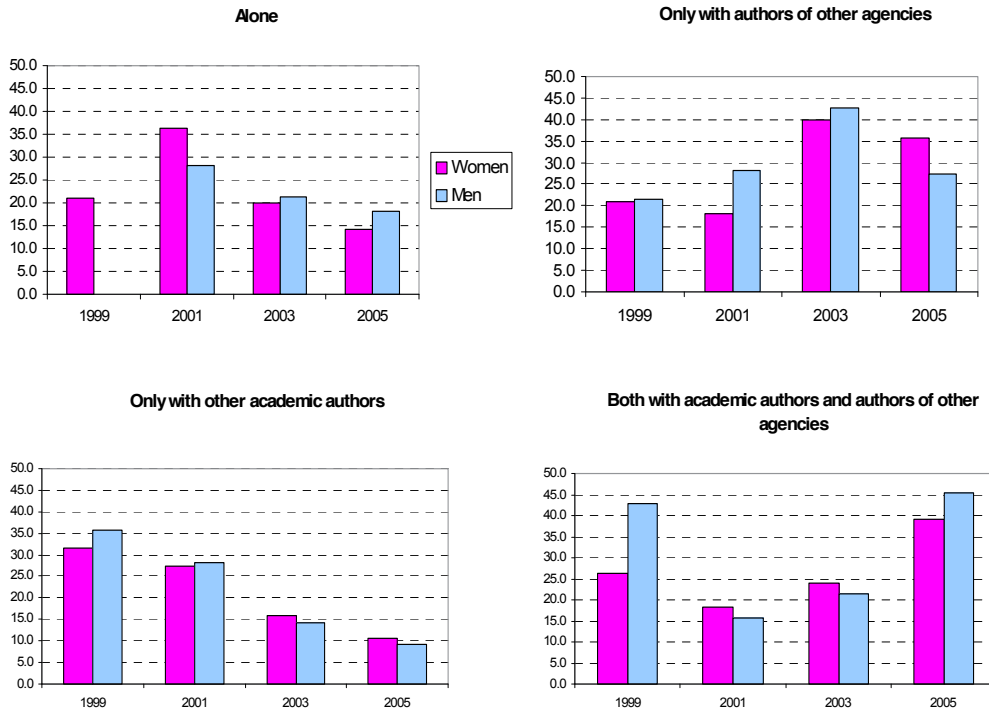
In general, the data represented in Figure 6 seems to reveal a sporadic collaboration between academics from various sectors referring to one single project presented in a determined conference of the GSP. Interdisciplinary collaborations seem to be sustained only when functional to one single project and don't seem to produce interdisciplinary cooperation over the long term. All things considered, this situation could be due to the fact that in various conferences, a high turnover among the participants can be observed, and therefore it is impossible, with the data available, to "follow" individual researchers over time.

Figure 6. Academic Authors by Disciplinary Sectors Other than Demography
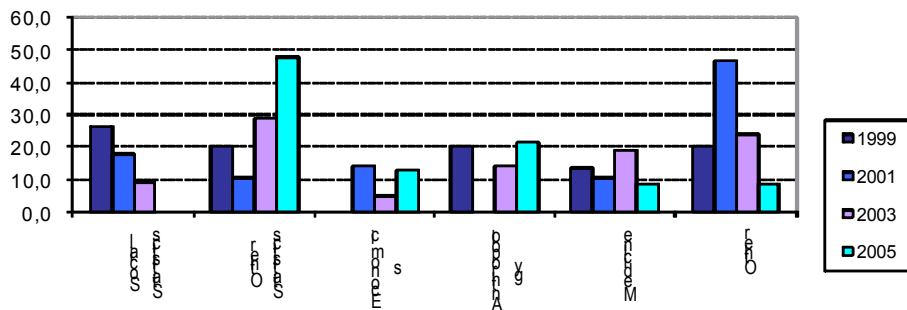
CONCLUSIONS

Using network analysis, we have studied how the Italian demographic community organizes its scientific relationships into collaboration patterns. We furthered the empirical analysis from the idea that presenting a co-authored paper in an official meeting represents a form of interaction among scientists that enriches the quality of the paper and the research process. These forums can stimulate the transmission of knowledge and experience from one author to another, avoiding self-referential works.

In this perspective, co-authoring a paper represents a possible form of interdisciplinarity apparent even when the authors, differing in their attributes and approaches to the proposed subject, belong to the same discipline. The quantitative database at our disposal allowed us to take into account other sciences (i.e. economics, sociology, social statistics, and anthropology) and other kind of institutions outside the university world, where population studies are diffused.

The empirical analysis of scientific interactions was necessary to set the original publication information into a relational framework. This represented a preliminary and fundamental step in calculating network analysis measures and to visualize the graphs. The books of abstracts are a collection of conference documents - while the data sets extracted from the abstracts were principally concerned with the authors and some of their characteristics: gender, affiliation, number of contributions, topic of the single paper, etc. This new way of reading and analyzing abstracts information allowed us to qualify every single author even taking into account his capacity to easily relate with researchers belonging to different areas of study.

The main results revealed an increasing participation at the GSP and a high turnover rate of participants over the four years observed. Other aspects showed fewer changes and proved to remain consistent over time. One example is the consistent predomination of academics in participation as well as productivity. Another element which was relatively consistent across years involved research groups. Research groups, if present, proved to be of small dimensions and to be homogenous in terms of gender and place of origin.

Although a high number of participants are represented by isolated nodes, the most common way of collaborating is the dyadic relationship and occurs most frequently between academics who are affiliated with the same university or with universities that are in close geographic proximity. In addition, an increase in women's participation can be globally observed with smaller research units showing greater female homogeneity.

According to the organization to which the author is associated, the type of collaboration tends to be constant. Opening up towards researchers of different areas is most common in the university world, even if such interdisciplinary collaborations appear to diversify themselves over time given the high turnover of the participants. Researchers of other research organizations are, however, more inclined towards collaborations within their affiliation (see sub-graph in 2001 network); they develop abundant collaborations due in large part to the presence of a small number of leading authors with the capacity to activate, organize, and maintain multiple collaborations (very centralized network).

The indications obtained on collaboration developed from the analysis in this paper, even if limited to the case of a specific convention, supply a rather clear picture regarding the way in which collaboration takes place in Italian Demography. Nevertheless, this database doesn't afford the possibility of following the scientific activity over time and, in particular, the strength of collaborations maintained by each author.

REFERENCES

Baldini R., Gallo G., Mancini P. (1998). La ricerca sulla Popolazione in Italia. Anni 1996-1997. Firenze: GCD.

Batagelj V., Mrvar A. (2003). Pajek - Analysis and Visualization of Large Networks. In: Jünger M., Mutzel P., (eds.). Graph Drawing Software. Berlin: Springer: 77-103.

Burch T. K. (2002). "Teaching Demography: ten principles and two rationales". *Genus*, LVIII, 3-4: 73-90.

Casacchia O., Mancini P. (1995). La ricerca sulla Popolazione in Italia. Anno 1993. Catanzaro: GCD.

Casacchia O., Gallo G., Mancini P (1996*).* La ricerca sulla Popolazione in Italia. Anni 1994-1995. Roma: GCD.

Chasteland J.C., Loriaux M., Roussel L. (2004). Démographie 2000. Une enquête internationale par Internet auprès des démographes. Louvain-la-Neuve: Academia-Bruyland: 367.

Farahat, H. (2002). "Authorship patterns in agricultural sciences in Egypt". *Scientometrics*, 55, 2: 157-170.

Gendrau F., Huix-Adamets S., (2003). Panorama institutionnel et scientifique de la recherche demographique dans le monde. Paris: CICRED.

Hoem J.M. (2007)."Demography, present and future", *MPIDR Working Paper*, 2007-019.

Liang L., Guo Y., Davis M. (2002). "Collaborative patterns and age structures in Chinese publications". *Scientometrics*, 54, 3: 473-489.

Liberman S., Wolf K.B. (1997). "The flow of knowledge: scientific contacts in formal meetings". *Social Networks*, 19: 271-283.

Melin G., Persson O. (1996). "Studying research collaboration using co-authorship". *Scientometrics*, 36, 3: 363-377.

Newman M. E. J. (2003). "Ego-centered networks and the ripple effect". *Social Networks*, 25: 83-95.

Rivellini G., Rizzi E. (2002). Science network in Italian population research: an application of Social Network Analysis. In: Atti della XLI Riunione Scientifica SIS. Padova: CLEUP: 215-218

Rivellini G., Rizzi E., Zaccarin S. (2006). "The science network in Italian population research: an analysis according to the social network perspective". *Scientometrics*, 67, 3: 407 – 418.

Rivellini G., Terzera L. (2008) "Partecipazione e contributi degli studiosi di popolazione ai più recenti convegni nazionali e internazionali delle associazioni scientifiche del settore". *Popolazione e Storia*, 1.

Russell Bernard H. (2005). Book Review. *Social Network* 27, 377-384.

Stoetzel J. (2006). "Sociology and demography". *Population*, 61, 1-2.

Tabutin D. (2005). Vers quelle(s) démographie(s)? Atouts, faiblesses et évolutions de la discipline depuis 50 ans. XXVe Congrès International de la Population, Tours.

Wasserman S., Faust K. (1994). Social Network Analysis. Methods and Application. USA: Cambridge University Press.